

In praise of open software

Freely available software, developed by researchers, is good for science and keeps commercial companies on their toes. In an era of quasi-monopolies, research institutions should encourage it.

Imagine how the web might look today had it been invented by Microsoft and made proprietary, rather than at the European Laboratory for Particle Physics (CERN), where it was made available free. Scientists tinkering with computers to create tools for their research for no profit have underpinned the computer revolution. The bounds of supercomputing are being pushed back by hugely demanding challenges, such as protein folding and the cosmos; many of the pioneers of the Internet are not Internet millionaires, but are still labouring in their laboratories.

The profit motive, and the investments that go with it, are often essential. The scrappy, early 'Mosaic' browser designed at the National Center for Supercomputing Applications at the University of Illinois only took off when some of the scientists who invented it went on to set up Netscape. But the abuse of commercial monopolies is also too evident, with much of the world having been held hostage to the dismal operating system DOS for more than a decade.

This issue — providing equitable access to all scientists and not just the richest — is about to become critical as companies rush to build bioinformatic tools for genomics. Tools that add value to genome data are to be welcomed, but as the licensing strategy being adopted by Celera Genomics becomes clear (see page 231), it gives new grounds for wariness. Unfortunately, restrictive material-transfer agreements are also becoming the rule even in publicly funded institutions (see page 236). While academic research centres are an important cradle for industrial development, it is crucial that the not-for-profit motive should be respected when the needs of research communities are best served in this way.

The high cost of some journals has attracted enormous attention over the past few years, whereas the high cost of software and the often

exorbitant licence charges have not. Most scientific software is proprietary, and beyond the reach of many poorer parts of the scientific community worldwide. All the more reason to be grateful, therefore, for the continuation of the open spirit in the tradition of Internet pioneers. Witness the group of Californian scientists developing sophisticated 'freeware' for DNA chip technology (see page 234). The software, which users say compares favourably with costly commercial software, can be downloaded from the web. Another example: scientists at the Max Planck Institute in Potsdam have made freely available a vast suite of plug-and-play tools, 'Cactus', that allows scientists from any discipline to use supercomputers without needing to know advanced computing techniques. A Japanese scientist is giving away E-Cell, a package that simulates basic cell processes. And so on.

The open-source movement has found its apogee in the Linux operating system developed by Linus Benedict Torvalds (see <http://www.cs.helsinki.fi/~torvalds>) as a 'hobby' — which IBM last week decided to put at the core of its hardware plans. Because the code is not proprietary, it is being built on and debugged by an army of amateur developers worldwide, many of them academic scientists.

In short, amateur software developers are playing a key role in keeping systems open. But such activities need to be encouraged and professionalized by academic institutions; plans in France to create a research centre to provide bioinformatic tools for industrial and academic researchers build on the tradition of the Centre d'Etude du Polymorphisme Humain, the US National Center for Biotechnology Information and the European Bioinformatics Institute. At a time when Microsoft looks as if it may be broken up (shades of AT&T) into 'Baby Bills', it would be ironic if science, and biology in particular, became a victim of new monopolies. ■

A step up for a few postdocs

One initiative in one Dutch institute, born out of necessity, provides an example to follow.

If there is one thing postdoctoral and contract researchers should know by now, it is that if they don't look after their collective interests, no one else will. Eminent academic panels and learned societies frequently speak up on their behalf, but can come to conflicting conclusions (see attempts in recent years by various panels to address the embarrassing surplus of postdocs in the life sciences in the United States). Research funding agencies and publicly funded employers seek to set new standards, but progress in improving status and rewards is slow and, ultimately, can only come by collective coherence, sustained campaigning and learning from the examples of successful initiatives. One such initiative may therefore be worth noting.

A survey of postdoctoral researchers by *Nature* and the European Science Foundation indicated that the Netherlands is a good place to be a postdoc as far as laboratory conditions are concerned (see *Nature* 397, 640–641; 1999). But a recent meeting of postdocs at a major Dutch research institute, the Netherlands Cancer Institute (NKI), highlighted their profound frustration at their lack of research independence and poor career progress and development. The fact that

postdocs took time out to collectively argue for better conditions, and that staff listened and acted, is unusual and commendable.

Some argue that the institute is simply fighting a rearguard action because of growing recruitment difficulties (see page 235). It is true that changes to employment law preventing institutes from awarding repeated contracts to an individual have strained the Dutch postdoc system and are galling to postdocs themselves. But postdocs often forget that what may seem expedient in the short term is not necessarily good in the longer term. A system of endless short-term contracts that lead nowhere may appear to offer a 'holding position' until a permanent job comes along, but, for most, such a system leads nowhere.

What more postdocs should be arguing for — a career structure — is just what some postdocs at the NKI asked for and are getting. Allowing postdocs independence by letting them apply for funding in their own right, and to take it with them if they move, is good for individual postdocs and good for the science. Other institutions should take up the challenge of providing for postdocs, and not simply wait until signs of discontent or serious recruitment problems develop. ■





Burn out?
NASA may end faulty
satellite's mission
prematurely
p232



Trade talks
Spectre of Seattle
looms over Montreal
negotiations
p233



DNA chips
Californian team
teaches researchers
self-sufficiency
p234



UK food safety
Krebs to become
head of post-BSE
standards agency
p235

Celera genome licensing terms spark concerns over 'monopoly'

Paris and Washington

Concerns that Celera Genomics Corporation may be poised to obtain a monopoly over human genome data are mounting in the academic research community, as the implications of the company's licensing policy on its soon-to-be-released sequence database begin to sink in.

In a statement earlier this month, Celera said: "When sequencing and scientific analysis of the human genome is completed, the consensus sequence data will be submitted for publication in a scientific journal. These published data will be made freely available to researchers around the world under a non-redistribution agreement."

Philip Green, a biocomputing expert at the University of Washington, argues that this "non-redistribution clause" represents a "significant departure from previous promises made by Celera that the sequence would be in the public domain; this presumably means, for example, that the data will not be submitted to GenBank."

Celera's president Craig Venter promised Congress under oath, in 1988 [http://www.house.gov/science/venter_06-17.htm], that data from the human genome would be made available to the public domain and that he would work with the National Library of Medicine, which runs GenBank. "Those statements implied to all observers that there would be truly no limitations on access or use," says one scientist working for the publicly funded international genome project.

Paul Gilman, director of policy and planning at Celera, confirmed to *Nature* that no data will be submitted to GenBank. Instead, sequence data will either be released on the company's own web site, or provided to researchers on DVD discs.

Gilman compares the terms of the proposed free public release with that of software licensing agreements. By opening the DVD package, the recipient would agree not to redistribute the data. "We do not want researchers to redistribute the data in competition with us," he says fearing that this could undercut Celera's business plan, which relies on selling subscriptions to its database to large pharmaceutical companies.



Lander: "All data should be broadly and freely available." Inset: Celera's mission statement.

But many researchers fear that the license will preclude free use of the data by scientists and that Celera may eventually close out all competitors. "I think it would be a disaster for biology if [a monopoly] happened to bioinformatics," says John Sulston.

Another genome scientist involved in the publicly funded project says that the licensing model adopted by Celera puts stringent limitations on use. "No other commercial database can use the Celera sequence to build a better annotated version," he says. "Their explicit intention is to obtain a monopoly on this field, driving out Incyte, Pangea, and a host of other creative companies whose ability to provide added value to the genome sequence is something we all want to flourish."

Other researchers are concerned that the agreement would prevent data contained on the DVD being used to design DNA chips for the whole human genome, or other genomic or proteomic experiments that would require use of the entire genome sequence. Gilman argues that Celera is not interested in selling raw data as such, but rather in licensing sophisticated software tools for its analysis. He defends the redistribution agreement. "People will have the entire completed genome from us before they have it from the public and they'll have it for free."

It is true that the Celera database, likely to be available this summer, will be the best available and very attractive to scientists in academia and industry. Although by late spring Celera will only have sequenced the genome to a depth of 4X (meaning that four bases of sequence have been generated for

every base of genome), its database will be close to the 10X gold standard — ironically, obtained only by merging its own 4X data with the public-domain 5X data. In contrast, the publicly funded project will only have its own 5X sequence to offer. "They can take our data, but we can't take theirs," points out one genome scientist. This lead will give Celera a serious competitive advantage in the period before the public project single-handedly reaches 10X coverage in around 2003. Gilman disputes that the Celera database will become less attractive with time: "It's not just about the data," he says.

Eric Lander, director of the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, says he "applauds efforts to add value to the genome" but adds "All data should be broadly and freely available. There is no proprietary genome." Lander predicts that dozens of companies will jostle to provide bioinformatic tools to analyse the human genome, and is concerned that the terms of Celera's licensing agreements may stifle this.

Celera is thinking of charging \$20,000 dollars annually per laboratory for basic access to its database, says Gilman, adding that universities with multiple laboratories would probably pay around \$5,000 per laboratory. Higher levels of access and a greater number of tools would probably cost more, he says. Academics could have "exactly the same access as the pharmaceutical companies, if they want it". Craig Venter adds: "If no one subscribes to it, we're out of business."

Gilman envisages that laboratories would subscribe to different suites of tools and access, with "deluxe", "economy" or "middle" packages. The company has not yet signed any contracts with academic laboratories, but is in discussion with some 15 universities, according to Gilman. "Our goal has always been to have academic subscribers."

Mark Magnuson, director of Transgenics/ES cell research at Vanderbilt University, says that the university is "interested" in subscribing to Celera's database, in particular to annotated data. But he says, "they're going to have to have information that's not in GenBank."

Declan Butler and Paul Smaglik

Merger of Glaxo Wellcome and SmithKline Beecham creates pharmaceutical giant

Munich

The merger of the British pharmaceutical companies Glaxo Wellcome and SmithKline Beecham, which comes into effect this summer, creates a giant which could rank number one in the world in terms of spending on research and development (R&D). It will boast a portfolio of 30 new drugs and 19 vaccines in clinical trial.

Glaxo Wellcome is currently the fifth-largest pharmaceutical company in the world in terms of turnover and SmithKline Beecham is ranked twelfth. The new company will have a stock-market value of £110 billion (US\$180 billion).

It displaces Aventis from the top position, a company created just a few weeks ago by the merger of Rhône-Poulenc and Hoechst. Meanwhile, the merger between Pfizer and Warner-Lambert, if it materializes, would create the second-largest company in the sector.

Savings made by merging the research sections of Glaxo Wellcome and SmithKline Beecham — estimated at \$415 million — will be reinvested in research, say the companies.

The new company plans to invest \$4 bil-

lion a year in research. Its nearest rival would be Pfizer Warner Lambert, whose component companies spent £2.7 billion in 1998.

Glaxo SmithKline will employ over 15,000 research staff. None of its research departments will be closed and there is no threat of significant redundancies amongst scientists, according to a press release. The companies also promise to maintain their scientific campuses at Research Triangle Park, North Carolina, and Philadelphia.

The merger will combine the power of Glaxo Wellcome's strength in combinatorial chemistry and SmithKline Beecham's strength in genomics and cell and molecular biology. Glaxo Wellcome stands to have its product portfolio rejuvenated through SmithKline Beecham's genomic assets.

The strategies of the companies, stressing genome research, are similar and their therapeutic target areas are complementary: in neurology, for example, Glaxo Wellcome is active in anti-migraine drugs and SmithKline Beecham in anti-depressants.

SmithKline Beecham was the first major pharmaceutical company to invest in

genomics, with its 1993 agreement with Human Genome Sciences, the commercial arm of the Institute for Genomics Research, where Craig Venter first made sequencing fashionable. As a result of its genomics activities, SmithKline Beecham now investigates between 150 and 200 new protein targets each year, compared with six to eight before 1993.

Glaxo Wellcome also has functional-genomics programmes and invests heavily in pharmacogenetics, a genomics-based activity that allows patients likely to benefit from a drug, or to suffer adverse side effects, to be identified before therapy is selected.

Given the long lead time for getting any new drug onto the market, analysts say it will be at least five years before it is clear whether the merged company will get a return on its investment in genomics.



Garnier: to head Glaxo SmithKline.

JOHNNY EGGITT/FT

NASA ponders termination of gamma-ray observatory

Washington

The US space agency NASA may retire the Compton Gamma Ray Observatory satellite as early as this March, unless it can ensure that the spacecraft will not crash back to Earth following last month's failure of an onboard gyroscope.

Launched in 1991, the \$600 million observatory has already outlived its planned lifetime, but was expected to stay in orbit for another eight to ten years.

Although the satellite can operate normally with its two remaining gyroscopes, it could not be pointed if another one fails. NASA has always intended to steer the spacecraft to a fiery re-entry over the ocean at the end of its useful life. An uncontrolled re-entry risks large chunks of

the 17-tonne observatory crashing into populated areas.

Engineers at the NASA Goddard Space Flight Center will know in mid-February whether the spacecraft can be controlled with one or no gyros. If not, other options include sending astronauts on the space shuttle to carry out repairs, boosting the observatory to a higher orbit or 'de-orbiting' the spacecraft as soon as possible, while it can still be controlled.

Compton has been a "very productive mission" and is still valuable, according to Kevin Hurley of the University of California at Berkeley, who chairs the observatory's scientific users' committee.

No other high-energy satellite can cover the entire sky; but the Compton mission has passed its peak of scientific productivity, and one gamma-ray telescope, EGRET, is barely operational. NASA's review of the project in 1998 called for it to continue at a reduced funding level, but said that: "Except for solar observations, the potential for new discoveries is diminishing."

There are several gamma-ray satellites in the works. The European Space Agency's International Gamma-Ray Astrophysics Laboratory, a collaboration with the United

States and Russia, is due to launch next year. NASA's SWIFT mission to hunt for gamma-ray bursts will begin in 2003. And HETE-II, a low-cost spacecraft run for NASA by the Massachusetts Institute of Technology, is due to reach orbit this spring.

HETE-II was supposed to be launched from Kwajalein Atoll in the Pacific Ocean on 28 January. But, following last month's failure of the Mars Polar Lander, NASA headquarters last week ordered HETE-II removed from its Pegasus launcher to undergo a final round of testing.

That sense of caution — and the desire to avoid embarrassment — will influence NASA's decision. If Goddard engineers fail to convince NASA headquarters that the spacecraft can be controlled without gyroscopes, an early de-orbit is likely. A space-shuttle rescue would be costly, and would take a long time to plan. Boosting the orbit would defer the plummet to Earth for a century or more.

Some defenders of Compton point out that it will not fall to Earth for at least six years. Another idea discussed — only half seriously, says Hurley — is letting the US defense department shoot down the satellite for target practice.

Tony Reichhardt



Down to Earth? Faulty gyroscopes may force NASA to steer Compton into the atmosphere.

NASA

Global-warming sceptics left out in the cold

Washington

A widely noted discrepancy between satellite-based and ground-based measurements of average global temperature does not call into question the fact that the Earth's atmosphere is warming up, according to a US National Research Council (NRC) panel.

The panel includes researchers who in the past have argued that the lack of evidence for warming from data gathered by satellite-based equipment cast doubt on global warming. Its unanimous conclusion therefore seems to refute one of the main arguments of opponents to the introduction in the United States of compulsory measures.

But positions on the global-warming issue are now so deeply entrenched that the finding is unlikely to move many policy-makers, according to congressional and administration staff on both sides of the debate. The Clinton administration has no plans to submit the Kyoto Protocol on greenhouse-gas emissions to the US Senate, and the Congress continues to oppose most actions or incentives aimed at reducing such emissions.

The NRC panel concluded that an increase in global mean surface temperature over the past 20 years "is undoubtedly real and is substantially greater than the average rate of warming during the twentieth century". Satellite data showing a lack of warming of the upper atmosphere were reliable, it said, but it rejected the contention that this called ground-based data into question. "We're saying emphatically that it is not valid to equate the two sets of measurements," explains John Wallace, chair of the panel and a climatologist at the University of Washington in Seattle.

The report states that the 20-year period monitored by satellite-mounted microwave sounding units was too short to indicate long-term climate behaviour. Wallace says that if earlier tropospheric temperature data from balloons are taken into account, and the record is examined over 30 or 40 years, the discrepancy between the surface and the troposphere disappears.

When the satellite-based data first appeared almost a decade ago, sceptics jumped on it as evidence that fears of global warming were overblown. "People started talking about the satellite data as a surrogate for the surface record," says Wallace. "There remains some residual misunderstanding that the two ought to be the same."

Wallace says that the disparity between the records may be due to short-term perturbations in the climate record — possibly caused by volcanic activity — or may point to limitations in existing climate models, which assume a close coupling between

mean temperatures at the surface and those in the upper atmosphere.

Sceptics of climate change have moved on from their earlier contention that the satellite data discredit the ground data to argue instead that the discrepancy between the two discredits the best available climate models. "What this tells me," says one republican congressional official who is sceptical about climate change and who was briefed on the report's contents last week, "is that the models need a lot of work."

John Christy, a panel member and atmospheric scientist at the University of Alabama at Huntsville, who specializes in analysing satellite-generated data, says the models "require

improvement", but adds that "it is not right to say that they are useless".

Indeed, few observers of the climate-change debate in Washington expect that the NRC's findings will have much immediate impact. "It's like a game of 'whack-a-mole'," says a democrat official on the Science Committee at the House of Representatives. "Every time you pin something down, another 'unresolved issue' comes up."

But Robert Park, a professor of physics at the University of Maryland and spokesperson for the American Physical Society, says the controversy is helping to generate "great science". The report is "evidence that the scientific process is working", he says. **Colin Macilwain**

Diversity convention in the balance

Washington

International negotiators gather this week in Montreal to try and negotiate a biosafety protocol that would regulate the international movement of living organisms. After the dramatic failure of last November's meeting of the World Trade Organization (WTO) in Seattle, hopes for agreement may well be equally forlorn.

Parties to the Convention on Biological Diversity (CBD) will meet to bridge divisions over whether a biosafety protocol should cover agricultural commodities, and whether it should take precedence over the rules on international trade set by the WTO.

As has become routine in environmental negotiations, deep divisions between the European Union and the United States and its food-exporting allies, including Canada and Argentina, will dominate the meeting. Europe favours a powerful biosafety protocol, whereas the US group's priority is to protect agricultural trade. Developing countries will broadly support the European position, and a small number of nations — known as the Compromise Group — will attempt to bridge the gap.

Only "a miracle" can produce an agreement, says Calestous Juma, former director of the CBD and now a professor at the Center for International Development at Harvard University. "The fundamental issue is between the European Union and North America: the rest of the world is coming along to watch."

Two main issues divide Europe and the United States, negotiators say. One is the application of the biosafety protocol to commodity crops: the European Union and developing countries want it to govern crop movement, while the United States wants an agreement covering only living organisms.

The second, more vexing question is



Up in smoke: the WTO meeting in November was dogged by riots and ended in failure.

whether the protocol should contain a 'savings clause', which would concede its subservience to WTO rules. "The European Union wants this to overrule the trading rules," says one European diplomat in Washington. "The United States will never agree to that."

Failure to break this impasse in Montreal — on top of two rounds of largely fruitless meetings since the parties failed to resolve the issue last February in Cartagena, Colombia — could spell trouble for the CBD, one of three international mechanisms established to protect the global environment at the 1992 Rio Earth Summit.

Juma says that failure to negotiate a biosafety protocol will damage the credibility of the convention as a means of tackling more complex issues, such as the sharing of genetic resources.

But a senior US government official says recent discussions have been "less ideological and more constructive" than at Cartagena, adding: "We come to Montreal prepared to negotiate in good faith."

Colin Macilwain

DIY microarrayers promise DNA chips with everything

San Francisco

DNA chips are one of the most exciting, powerful — and expensive — genome technologies. A group of Californian scientists has now developed a do-it-yourself kit, available on the web, designed to bring them within the reach of laboratories on tight budgets.

DNA chips, or microarrays, consist of wafers, usually made of glass, etched with hundreds of thousands of microscopic wells. Each well contains a short stretch of DNA, which bonds with any matching strands in a sample, the matches being identified using fluorescent tags. This high-throughput screen is ideal for detecting specific genes, and has applications from evolutionary biology to the diagnosis of disease.

Two years ago, a group of scientists concerned about the cost of microarray chips and chip-making equipment set up a website with instructions on how to build a microarray system for making custom chips (<http://cmgm.stanford.edu/pbrown/mguide/>). The group includes Joseph DeRisi of the University of California at San Francisco, Mike Eisen of Lawrence Berkeley National Laboratory, and Pat Brown, Paul Spellman and Max Diehn, all of Stanford University. DeRisi is a former postdoc in Brown's laboratory, where the system originated.

A 'freeware' database designed for microarray data is the team's latest offering. The software, which was posted to the group's new website on 1 January, allows users to store, sort and analyse genetic data (see <http://www.microarrays.org>).

Janet Hager, a cell biologist at Yale University, says that comparable commercial software would cost up to \$200,000.

The "Mguide", as the "Brown lab's complete guide to microarraying" is called, leads a researcher through buying, assembling and operating a microarray chip maker. The main equipment is a computer-controlled robotic arm — originally designed for the semiconductor industry — that drops DNA specimens onto chips. Chips are then analysed using a scanner, a separate piece of equipment that can be made or bought.

Using the website, a researcher can build a microarrayer for about \$30,000 — half the price of a commercial machine, whose costs have themselves dropped substantially in recent years. It costs about \$30,000 to buy a scanner and slightly less to build one.

Kenneth Burtis, a molecular biologist at the University of California, Davis, who has built a microarray from Mguide, says it helps to be a "tinkering" person. DeRisi, a fellow in biochemistry and biophysics, says: "A 14-



Home cooking: building your own microarrayer costs half as much as buying one. Inset: DeRisi

year-old could build one in about ten hours." Home-made microarrayers are in laboratories at institutions such as Yale University, Uppsala University in Sweden and the Institute of Food Research in Norwich, England.

Scientists from these institutions took part in a microarray course last November at Cold Spring Harbor Laboratory in New York, where Brown, DeRisi and colleagues provided intense instruction. The 16 participants — out of 125 applicants — made four microarrays and learned how to use them.

Cold Spring Harbor is planning a second course in June. This is planned to concentrate on the use of microarrays. Since the first course, the Mguide website has been upgraded with a system that greatly expands the microarrays' production capability.

Hager took a microarrayer back to Yale, where scientists from 25 research projects have teamed up to use the equipment, funded by a grant from the National Cancer Institute. It "was a challenge to build, but rewarding," says Hager. "The challenge for the future will be bioinformatics, interpreting the vast amount of data" from the microarrays.

Burtis says that the group deserves praise for bucking what he describes as a trend towards commercialization in bioinformatics, by sharing its knowledge without seeking to make a profit.

Affymetrix, a company based in Santa Clara, California, and the leader in DNA-chip technology, sells its chips for \$100 to \$2,000 each. Dave Crawford, marketing director at Affymetrix, says the scientists are "providing a valuable service to the academic community" and have "contributed a tremendous amount" to the microarray market. Affymetrix will introduce a range of cheaper chip products targeted at academic researchers later this year, he says.

Agilent Technologies, Motorola, Corning and other companies are following Affymetrix into the microarray market, and competition is expected to force down prices.

Rex Dalton

Australian university chiefs attack plans for research funding

Sydney

The Australian Vice-Chancellors' Committee has criticized a white paper released by the education minister David Kemp as being "flawed" by the government's refusal to accept that Australia's research base needs major additional investment.

Kemp says that the white paper, which seeks to tie research to its commercial potential, is intended to make Australia competitive in "the global knowledge economy". Kemp also confirmed legislation to establish the grant-giving Australian Research Council (ARC) as "an independent body with broader membership and a more strategic role".

Criticism has focused on the government's refusal to restore cuts to research budgets. Deryck Schreuder, the acting president of the vice-chancellor's committee, says that the "key contention" that Australian funding is comparable to that of other countries in the Organisation for Economic Cooperation and Development is based on outdated figures and shows a complacency that has left university funding "critical".

Kemp has jettisoned six of the more contentious aspects of his proposals (see *Nature* 402, 113; 1999), notably a voucher scheme for postgraduate students, a shorter completion time for graduate degrees, and the abolition of a block-grants programme for research infrastructure.

The Federation of Australian Scientific and Technological Societies is also challenging the government's statistics. Its new president, Sue Serjeantson, wants an independent audit of public spending on research "that permits valid international comparisons". Serjeantson claims the government has inflated its figures by including grants for the humanities, economics and social sciences.

Deane Terrell, vice-chancellor of the Australian National University, welcomed "the emphasis on performance-based funding for research training and infrastructure," but says that the problem of more money "remains to be addressed".

The Australian Academy of Science welcomed the strengthening of the ARC, but attacked the use of "simplistic formulae" to allocate money. It advocated an assessment exercise that would reward excellence.

Peter Pockley

Dutch institute forced to respond to crisis in recruitment of postdoctoral researchers

Hoeven, Netherlands

The Netherlands is faced with a crisis in the recruitment of postdoctoral researchers, with scientists increasingly reluctant to accept living on a series of short-term posts.

Figures from the medical council of the country's largest public research agency, the Netherlands Organization for Scientific Research (NWO), show that last year it filled less than a fifth of postdoctoral positions within its programme grants (see Table), only a third in its research support grants and just two thirds of its personal fellowships. Edvard Beem, deputy director of health research at the NWO, says that as a result it has been impossible for the council to request additional research funds from the NWO.

Problems in recruitment have also surfaced over the past two years at the Netherlands Cancer Institute (NKI) — a leading national research institute. Staff there say it is difficult to retain and attract postdocs to a country that does not have a large critical mass of research work.

Although there is no shortage of postdoctoral fellowships, the number of tenured positions in academia is limiting. Senior postdocs in particular feel unable to continue their scientific careers in the Netherlands



Peters: postdocs want permanent positions.

as short-term fellowships are unattractive to postdocs looking for career advancement and tenure. NKI staff and postdocs have dubbed this situation the 'postdoc paradox'.

"While universities and institutes are looking for graduate students and postdocs, there is, at the same time, a population of ageing postdocs looking for a permanent position and recycling through standard postdoc positions," says Peter Peters, a researcher at NKI.

Unsurprisingly, 90 per cent of the postdocs are ultimately looking to find permanent university or research positions. "Postdocs that don't make it have limited alternative prospects in the Netherlands because of limited career possibilities in biotechnology and pharmaceutical industries in the country, and the general decay in existing research jobs in industry," explains Peters.

Recruitment difficulties are due to a variety of factors, he continues, most importantly a lack of career development and

prospects. Some universities have had to re-advertise postdoc jobs with better salaries and contract terms in order to recruit.

NKI's new director — Anton Berns — last week responded to postdoc concerns, by establishing a novel junior faculty position that would give postdocs some degree of independence and career prospects. Junior faculty could apply for project grants with a senior staff member, would supervise graduate students, serve as senior authors on papers, carry projects with them when relocating to another laboratory and would also have their salaries boosted by about \$5,000.

Such a change is timely within a national, political context. A government report on science funding, published last year, stated that "the research establishment must itself take responsibility for developing a challenging personnel policy". In response, NWO will make available \$0.7 million per annum to improve the prospects for women in research and the science ministry will contribute a further \$0.25 million. **Natasha Loder**

Influx of personnel onto medical council grants

Type of grant	Granted (Jan 99)	Filled (Dec 99)	%
Fellowship support	12	8	66
Research support	40	12	30
Research programme	71	13	18

Appointment of UK food standards chief under fire

London

Zoologist John Krebs, a former chief executive of the UK Natural Environment Research Council, has been appointed head of the country's Food Standards Agency (FSA), created in the wake of the crisis over bovine spongiform encephalopathy (BSE).

His nomination has been criticized by the powerful UK Consumers' Association, which expressed its "disappointment" that the government had not appointed a "strong, credible consumer chair". It argued that this would have been more in line with the food agency's goal of providing impartial assessments based on the best available scientific advice.

But former colleagues of Krebs are unanimous in their opinion that his diplomacy, fairness and ability to take tough decisions make him a good choice to head the FSA. One observer says, "he is not beholden to either side".

Krebs says his top priorities in the coming weeks will be to meet the representatives of consumer groups and industry to understand "more about what

the issues in food safety and standards are". Krebs describes his job as being "at the interface between science and policy — a tremendously important and exciting area. Decisions [of the FSA] will have a firm basis in science, but science is not the total story".

Mistrust of public experts is rife since the BSE crisis (see *Nature* **400**, 389; 1999). Krebs says the agency will try to rebuild trust by publishing the advice it gives. "If we make a recommendation on a particular issue and explain publicly why we've come to the conclusion and made the recommendation, and if ministers then choose to go a different way, then the public would quite rightly ask why," he says.

Krebs has a reputation for taking accurate but unpopular decisions. He was responsible for



Krebs: trying to rebuild trust.

of whether badgers with tuberculosis pass the disease to cattle. The experiment, which included the culling of around 12,500 badgers over five years, faced strong opposition from conservation and animal-welfare groups (see *Nature* **394**, 821; 1998).

Until last September, Krebs was the chief executive of the Natural Environment Research Council. In 1988 he was appointed Royal Society research professor in the department of zoology at the University of Oxford.

In the light of the decision to appoint Krebs, the Consumers' Association is calling for the remaining FSA board members to have a stronger consumer focus. The association says it wants an end to the conflict of interest at the Ministry of Food, Fisheries and Agriculture between protecting consumers and promoting the food industry.

Krebs is certainly expected to do well in terms of the relationship between the FSA and the food ministry. He is also on close terms with the government's chief scientific adviser, Robert May.

N. L.

Mouse geneticists call for unified rules of exchange

Munich

Mouse researchers face a maze of Material Transfer Agreements (MTA) governing the exchange of research tools. A group of mouse geneticists has suggested that funding agencies should agree on a standardized, international MTA that would reduce paperwork and maximize free exchange.

The proposal comes from a task force of the International Mammalian Genome Society. It was prompted by concern about restrictions placed on the use of research tools, particularly by biotechnology companies.

Scientists say that an increasing amount of time is spent analysing MTAs. But while many publicly funded research organizations are sympathetic to the goal of standardizing MTAs, they are sceptical as to whether the proposed contracts — one for exchange between academics and one for exchange between academics and industry — would be universally applicable.

"Everyone wants to see free exchange of research tools," says Steve Brown, director of the UK Medical Research Council's Mouse Genome Centre in Harwell and a member of the task force, which was set up in 1997. "But it will not happen overnight."

"We are not in a crisis," he says. "Ninety-nine per cent of the huge number of research tools that biologists exchange are problem-free. But the issue needs to be addressed, because the one per cent of problems tend to have far-reaching consequences for basic research — for example, the *Cre-lox* issue."

The *Cre-lox* case led to the setting up of the task force. *Cre-lox* technology allows target genes to be removed from specific cells. It was patented by the pharmaceutical company DuPont, which demanded license fees and 'reach-through' rights to the profits from future discoveries made with the technology.

DuPont has agreed with the US National Institutes of Health (NIH) to allow its researchers to use *Cre-lox* technology without paying license fees (*Nature* 394, 819; 1998). But this agreement does not extend to European researchers.

As well as companies, academic institutions often demand strict MTAs. Many research bodies require first rights to negotiate a license for any invention that arises from the use of exchanged research tools.

The increasing number of MTAs "is a big administrative burden," says Philip Avner, of the Pasteur Institute's Unit of Molecular Mouse Genetics in Paris. Martin Hrabe de Angelis, head of the mouse-mutagenesis programme at the GSF, a national research centre in Munich, says that discussions with legal departments are "a waste of time." Stan-

dard rules would help researchers deal with their legal departments, which tend to want strict MTAs, he says.

Unforeseen problems pop up regularly. For example, the GSF recently refused to transfer its mutant mice to NIH researchers, because the NIH insisted that the GSF accept full liability under US law for any damages resulting from the use of the mice.

The NIH has published guidelines for more relaxed and standardized MTAs (see *Nature* 403, 10; 2000) — again partly because of the *Cre-lox* controversy — which do not have such strict indemnity clauses. Its proposed standard form makes no reach-through demands, and does not require for results to be seen before publication.

But the guidelines are only intended to cover situations with no obvious commercial implications, says one NIH official. Transfer of materials "cannot be solved on a global basis," and a standard form cannot broadly apply.

A spokesman for the MRC, which issues relatively unrestrictive MTAs between academics, agrees. "It is doubtful that a single form could serve the many different circumstances." Axel Polack of the GSF's legal department adds that the centre might be reluctant to adopt the standard form proposed by the task force: "our interests would be unlikely to be protected".

The GSF and other research centres in Germany are discussing the possibility of a single MTA strategy with the Max Planck and Fraunhofer societies, which run basic and applied research institutes, respectively.

But the participants do not see eye-to-eye. A typical MTA issued by the Max Planck Society requires only that it be kept informed of research results. The GSF requires that a recipient submits any results before publication, so that it can screen these for intellectual property protection and obtain first rights to negotiate a license.

Alison Abbott



Business lobby set to take EPA to court over data access

Washington

Controversial rules giving greater access to scientists' raw data under the US Freedom of Information Act (FOIA) are to be tested, following a request from the US Chamber of Commerce to see data held by the Environmental Protection Agency (EPA).

The chamber has submitted three requests for the data from university studies on the sulphur content of petrol, the health effects of particulate matter and the vulnerability of poor people to pollution. The EPA used these data to justify environmental regulations.

The requests exceed the scope of the rules laid down last year by the White House Office of Management and Budget (OMB) (see *Nature* 401, 732; 1999). As a result, the EPA will probably refuse to release the information. But the Chamber of Commerce says that in that case it will go to court to get the data.

A court showdown would bring simmering tensions in the scientific community to the boil. Scientists have opposed FOIA access to university researchers' data, arguing that it would lead to the harassment of researchers in sensitive fields. The business community says it needs such access to challenge what it views as unreasonable regulation by the EPA and other agencies.

Harvard University's refusal to hand over data from a study on particulates led Senator Richard Shelby (Republican, Alabama) to introduce a 1998 amendment stating that "all data" from university research performed for the government should be accessible under the FOIA (see *Nature* 397, 459; 1999).

The OMB was ordered to draw up rules to implement the amendment. But the Chamber of Commerce has dismissed these as "half-hearted".

The OMB rules exempt certain classes of data and only apply to research that underpins new regulations. The Chamber of Commerce is requesting all of the data from studies that underpin old regulations.

If refused, the Chamber will argue in court that the OMB rules conflict with the intent of both the amendment and the Freedom of Information Act itself.

University scientists are concerned that FOIA requests will be costly and may impugn their independence. The US Chamber of Commerce argues that publicly funded work should be available to the public.

Colin Macilwain

NIH under fire over gene-therapy trials...

Washington

A leading US congressional Democrat last week attacked the National Institutes of Health (NIH) for its failure to properly oversee patient safety in gene-therapy clinical trials, and suggested that responsibility for this be transferred to the Office of the Secretary of Health and Human Services (HHS).

Congressman Henry Waxman (Democrat, Los Angeles) wrote to Ruth Kirschstein, the acting director of the NIH, on 10 January, saying that he was "dismayed" at the frequency with which NIH-funded gene therapists failed to report serious adverse events to the agency.

Waxman referred to figures in a letter he received last month from Harold Varmus, then the outgoing NIH director. Varmus reported that 652 of 691 serious adverse events in human trials using adenoviral vectors were not immediately reported to the NIH, as required by its regulations — a 94 per cent non-compliance rate.

The figures were generated by an NIH request to scientists last October for reports on serious adverse events, after an 18-year-old research subject died in a trial using adenovirus at the University of Pennsylvania. Investigators responded with reports of 691 events, only 39 of which were reported at the time they occurred.

Waxman, who will chair the House Government Reform and Oversight Committee if the Democrats re-take the House of Representatives this November, asked Kirschstein to comment by 21 January on this failure.

Anne Thomas, a spokeswoman for Kirschstein, claims that the NIH has "served the public interest very well" by helping to ensure the safety and efficacy of gene therapy. But, she added, "there is always room for improvement".

Waxman implied that the NIH bore some responsibility for investigators' under-reporting by having failed to give clear instructions to its Recombinant DNA Advisory Committee (RAC). The committee publicly oversees gene-therapy protocols and is the NIH body that receives reports of adverse events from scientists. Reports must also go to the US Food and Drug Administration (FDA).

"I am particularly interested in whether you agree that the NIH has contributed to... non-compliance by raising doubts concerning the RAC's existence [and] by curtailing the RAC's authority over gene-therapy protocols," he wrote.

Waxman opposed a 1996 proposal by Harold Varmus, then the NIH director, to disband the RAC (see *Nature* 382, 197; 1996). The proposal was later modified to



Waxman: "dismayed" by reporting failures.

preserve the committee, but without its former authority to approve or disapprove particular experiments. That role now rests solely with the FDA, which by law must keep its deliberations private. Some people in industry say that investigator non-compliance is beside the point, because the NIH does not have the legal authority to require reports of adverse events — this authority rests with the FDA — or the human resources to assess them.

"The question of compliance is misconstrued," says Barrie Carter, the director of research and development at Targeted Genetics, a Seattle gene-therapy company. "Even if the data are sent to [the NIH] it's not clear they can compile them." He doesn't think that moving the NIH's gene-therapy oversight to the HHS secretary's office is useful.

Waxman's letter also laments NIH's failure to establish a public database of human

gene-therapy protocols, including reports of adverse events. The agency announced plans for the database in 1996. He asks the NIH to have the database operational by the end of this year.

The NIH "must remedy the apparent problems in its oversight of this important field of research", wrote Waxman. He then asked Kirschstein to comment "on whether the public interest would be best served by transferring the RAC and [its administrative office] from NIH to the office of the HHS Secretary".

According to Michael Werner, the Biotechnology Industry Organization's lobbyist on gene-therapy issues, the role of the RAC "is not to be a regulatory agency, it's to do public discussion of social and ethical points". He adds that "whether or not that takes place in the office of the [HHS] secretary or at NIH may be beside the point".

Moving the RAC to the higher-profile HHS secretary's office would mirror a similar move already underway — the NIH office that oversees research involving human subjects was recently elevated to the office of the HHS secretary.

Meredith Wadman

...as panel seeks help for trial host nations

Washington

The US National Bioethics Advisory Commission may recommend that protocols for US clinical trials carried out in developing countries should be adapted to the local health infrastructure so the host country can gain maximum benefit from the trials.

The commission is expected to issue guidelines this summer on US researchers' obligations to host developing countries and to their research subjects. A panel of the commission met last week to discuss the extent to which the United States should assist host countries. Proposed involvement ranged from tailoring trials to a developing nation's biomedical capacity to building up the country's infrastructure to match the needs of the research.

"It does no good to develop therapies that the host country cannot continue to use," Robert Levine, a professor at Yale University

School of Medicine, told the panel. For example, a trial in sub-Saharan Africa on how to minimize the spread of HIV from pregnant mothers to their fetuses would seem a worthy effort in the short term, but would not necessarily be of long-term help to the community.

If the community wished to continue the research after the US clinicians had left, it would need prenatal facilities to allow doctors to identify HIV-positive mothers and the equipment and trained personnel necessary to administer the anti-AIDS drug AZT intravenously. It would also need ample supplies of baby formula and clean water to provide an alternative to breastfeeding, as infants would otherwise face an even greater health risk from diarrhoea.

To justify carrying out clinical research in developing countries, US research goals should match the health needs of the host country and should

provide the country with the "highest attainable and sustainable" benefits possible after the end of the trial, Levine suggested. For example, a trial of an AIDS vaccine in a developing nation should include provision for further distribution of the vaccine should the trial prove successful.

Ruth Faden, director of the Bioethics Institute at Johns Hopkins University, told the panel that scientists change the community where the studies take place merely by setting up clinical trials, as they introduce — sometimes temporarily — higher levels of health care. As a result, they have an obligation to help that community after the trial ends. "You can't leave the place the way you found it," said Faden, adding that the US government should provide aid to build up infrastructure in developing nations where US-sponsored research takes place.

Paul Smaglik

Tougher rules will put US farmers off planting GM corn

Washington The US Environmental Protection Agency (EPA) has announced tight new controls on the planting of *Bt* corn, which is genetically modified to produce its own insecticide. The rules come into effect immediately, and require farmers to plant larger areas of unmodified corn nearby to limit the modified corn's ecological impact.

The rules are likely to hasten the decline in the acreage of modified corn grown in the United States this year. Surveys show that farmers fear low export demand for genetically modified foods.

The EPA, which is responsible for the regulation of agriculture's impact on the environment, designed the rules to try and prevent the evolution of insect resistance to *Bt* insecticide, and to reduce the impact of *Bt* corn on wildlife.

EU research chief seeks a debate on strategy

Munich European Union research commissioner Philippe Busquin this week launched a debate on research strategy in Europe. At present, research is too fragmented, he says, and

there is little coordination between research policies in the union's 15 member states.

Busquin resolves to convert into action six months of discussion with European institutions, including the Council and the Parliament, as well as with the scientific community. He stresses, however, that the Commission will play a primarily catalytic role, as most problems can only be addressed at the national level.

European science's many woes have frequently been aired but never solved. These include few women in senior research positions, too little mobility of researchers and inadequate networking of centres of excellence.

Kennewick Man is a Native American

Washington The controversial human remains known as 'Kennewick Man' (*Nature* 397, 551; 1999) are estimated to be 9,400 years old, according to radiocarbon dating of two small bone samples.

The tests, done last September, confirm earlier estimates. The US Department of the Interior said the result proved that the bones are Native American, and therefore come under the protection of the Native American Graves Protection and Repatriation Act.

Although scientists want to study the

samples, five tribes in the American Northwest have claimed them as their own, and hope to re-bury them. The next, more difficult step is to determine Kennewick Man's 'cultural affiliation'. DNA testing may not be possible, according to the Interior Department, which is under court order to settle the matter by 24 March.

Celera buys into Chinese biotech company

Paris The US genomics company Celera has bought 47.5% of Shanghai GeneCore BioTechnologies Ltd. The stock was formerly held by Axys Pharmaceuticals Inc., a Chinese genomics company active in the country's academic and industrial centres. Another branch of Celera's parent company, Perkin Elmer Biosystems, also owns 47.5% of Shanghai GeneCore.

Craig Venter, Celera's president, said that the acquisition was part of the company's global expansion plans, and would provide it with access to plant, animal and human genetic diversity in China. "Access to this diversity will be fundamental in Celera expanding its genomic information," he said. Ming-Wei Wang, chairman and general manager of Shanghai GeneCore said that access to Celera's data and expertise would boost genomics in Asia.

Automated patent database launched

London The World Intellectual Property Organization (WIPO) has begun work on a worldwide, automatic patent database with a consortium of information technology companies.

The project, which has a budget of 40 million Swiss francs (US\$25 million), is called IMPACT — Information Management for the Patent Cooperation Treaty. This treaty allows for a single international patent application valid in any of 106 states. An automated system is needed because of the rapid expansion in international patent applications handled by WIPO. These rose from 2,625 in 1979 to over 70,000 in 1999.

WIPO say that the database will benefit the public by making patent information available in electronic form. The project is expected to be complete in 3–4 years.

UK rheumatism researchers carry off Crafoord prize

London Ravinder Maini and Marc Feldmann, of the Kennedy Institute of Rheumatology in London, have won the Crafoord Prize for their work on the mechanisms of pathogenesis in rheumatoid arthritis.

Maini and Feldmann worked on the importance of cytokines — signal molecules — in rheumatoid arthritis, and identified the signal molecule TNF- α as a therapeutic target. Newly developed drugs block TNF- α .

The prize will be awarded by the Royal Swedish Academy of Sciences this September, and includes US\$500,000, to be split between the winners.

Genomics institute planned near Paris

Paris French research agencies are involved in discussions aimed at creating an IMAGE Institute for Integrated Genomics and Biotechnology at Evry near Paris. The existing public IMAGE consortium (Integrated Molecular Analysis of Genomes and their Expression; <http://www-bio.llnl.gov/image/>) — created to provide industrial and academic researchers with research tools — is in discussions with the Centre National de la Recherche Scientifique, other research agencies and the science ministry.

The group, which has collaborated with Synthelabo-Sanofi, Genome Express, and the Amersham Pharmacia Biotech MTAP consortium, and includes over 30 pharmaceutical, agrochemical and biotechnology companies, is expected to announce a major deal with a leading

bioinformatics firm that would underpin the creation of the centre of excellence in computational genomics.

“Such partnerships are based on balanced management of intellectual property rights. With our partners, we focus on valuable intellectual property rights, and we do not believe that applying blindly for patents on everything is meaningful,” says genome scientist Charles Auffray.

Not-for-profit publishers aim to win academics over

Paris The US Association of Research Libraries and the Scholarly Publishing and Academic Resources Coalition (SPARC) are launching a campaign, called ‘Create Change’, among academics to promote grassroots support for not-for-profit scholarly publishing.

The internet means that scholarly publishing is at a “turning point,” says Rick Johnson, SPARC Enterprise Director. There is “a historic opportunity to build a system that responds to [scholarly] needs rather than largely commercial interests. With this campaign, we’ll expand the discussion and highlight the choices available.”

The campaign, prompted in particular by steep inflation in journal prices, will include a series of brochures, web resources, advocacy tools and a speakers bureau.

Local data are vital to worldwide conservation

Sir—Red Data Books and Red Lists of threatened species have drawn attention to species at risk of extinction worldwide^{1,2}. As the number of national Red Lists increases, more information becomes available to assess the global status of species. Unfortunately, global assessments coordinated by the World Conservation Union (IUCN) are not making full use of this wealth of local data, often giving a misleading picture of the status of a taxon.

In our study of this problem, we excluded widely distributed threatened species, whose risk of extinction may vary from country to country, and focused on nationally endemic taxa, where national and global Red Lists should be identical. We found that national assessments tend to incorporate data from global assessments, but the reverse is much less frequent. This decreases the efficiency of conservation at the national level, where actions are most likely to have an impact; local organizations have difficulty raising international funds because donors believe, wrongly, that international lists are more accurate.

We compared global and national lists of threatened endemic species for Argentina³, Bolivia⁴, Ecuador⁵ and Venezuela⁶. Although most South American countries have national Red Lists⁷, only these four use the new IUCN Red List categories used for global assessments⁸. The combined national and global lists cover 173 taxa (Table 1). All taxa should be included in both lists; in fact, only a quarter are. Over half of the taxa listed nationally are not listed globally, while more than three-quarters of the taxa included in the global lists also feature in national lists. Of those taxa included in both lists, the categories they are assigned to in each list are more likely to differ than not ($\chi^2 = 4.22$, d.f. = 1, $P < 0.05$). Agreement between national and global lists is best for birds, followed by mammals and reptiles (Table 1).

We suggest that these disparities are due to a lack of communication between

national assessments based on geography and global assessments based on taxonomy. National Red Lists are usually produced by local organizations. In contrast, IUCN compiles global Red Lists through its Species Survival Commission (SSC), a network of more than 5,000 volunteer scientists organized into areas of taxonomic interest.

To improve the flow of information between national and global Red Lists, data from geographically based lists must be compiled, cross-referenced, and made available to global assessors. To this end, the SSC should create an interdisciplinary group to administer a website containing the national Red Lists of the world. By standardizing the process of submitting information to the website, a 'virtual' specialist group would be both cost-effective and relatively easy to manage; it would significantly improve the quality of Red Lists at both levels and would make research and conservation more effective. **Jon Paul Rodríguez***†, **Gillian Ashenfelter***, **Franklin Rojas-Suárez‡**, **Juan Javier García Fernández§**, **Luis Suárez¶**, **Andrew P. Dobson***

*Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544-1003, USA

†Centro de Ecología, Instituto Venezolano de Investigaciones Científicas, Apartado 21827, Caracas 1020-A, Venezuela

‡PROVITA, Apartado 47552, Caracas 1041-A, Venezuela

§FUCEMA, Pringles 10, 3P, 1183 Buenos Aires, Argentina

¶EcoCiencia, Isla San Cristobal 1523 e Isla Seymour, Casilla Postal 17-12-257, Quito, Ecuador

1. Scott, P., Burton, J. A. & Fitter, R. in *The Road to Extinction* (eds Fitter, R. & Fitter, M.) 1–6 (IUCN/UNEP, Gland, 1987).
2. Collar, N. J. *Oryx* **30**, 121–130 (1996).
3. García Fernández, J. J., Ojeda, R. A., Fraga, R. M., Díaz, G. B. & Baigún, R. J. *Mamíferos y Aves Amenazados de la Argentina* (Fundación para la Conservación de las Especies y el Medio Ambiente, Buenos Aires, 1997).
4. Ergueta, P. & de Morales, C. (eds) *Libro Rojo de los Vertebrados de Bolivia* (Centro de Datos para la Conservación, La Paz, 1996).
5. Granizo, T. M. et al. in *Lista de Aves de Extinción en el Ecuador* 1–31 (IUCN-Sur, CECIA, INEFAN, EcoCiencia y BirdLife International, Quito, 1997).
6. Rodríguez, J. P. & Rojas-Suárez, F. *Libro Rojo de la Fauna Venezolana* (PROVITA, Fundación Polar, Caracas, 1995).
7. WCMC *Biodiversity Data Sourcebook* (World Conservation Press, Cambridge, UK, 1994).
8. IUCN 1996 *IUCN Red List of Threatened Animals* (IUCN, Gland, 1996).

Table 1 **Nationally endemic South American animals in national and global Red Lists**

Taxon	National lists		Global lists		Both lists	
	Also in global list		Also in national list		Taxa simultaneously	
	Number	(%)	Number	(%)	Number	in both (%)
Mammals	29	69	27	74	36	56
Birds	66	76	58	86	74	68
Reptiles	4	50	3	67	5	40
Amphibians	9	0	0	0	9	0
Bony fishes	8	0	4	0	12	0
Crustaceans	6	0	1	0	6	0
Insects	31	3	1	100	31	3
All taxa combined	152	46	94	78	173	42

There's a place for the theory of everything

Sir—As the originator of the term 'Theory of Everything' (TOE) to represent the ambition of string theory to describe all the elementary particles and their fundamental interactions (*Nature* **323**, 595; 1986), I was dismayed by the tone of a recent review by my old friend George Ellis of *The Quest for Unity* (*Nature* **401**, 527; 1999) by Etienne Klein and Marc Lachièze-Ray.

My definition of a TOE — a theory to "unify all the fundamental interactions" and "explain the number and couplings of all the elementary particles" — is not different from Ellis's "theory unifying all fundamental forces into one", but I take issue with his (and your headline writer's) assertion that this "is ... irrelevant to most physicists".

In addition to the cosmologists cited by Ellis, the TOE ambition of particle physics is relevant to any astrophysicists interested in topics such as black holes, gravitational waves, active galactic nuclei, gamma-ray bursters, dark matter, structure formation, supernovae, high-energy cosmic rays and so on. Moreover, the theoretical tools of quantum field theory, conformal field theory, topological solitons and so on, discovered or developed by particle and string theorists, are central to much of condensed-matter theory, and even of interest to pure mathematicians; witness the Fields medal awarded some years ago to Edward Witten, even before the advent of Seiberg–Witten theory.

Like Ellis, I lament the fact that physics "is becoming ever more fragmented into disparate subjects", but particle physicists are hardly to blame for the richness of the phenomena (relativity, quantum theory and so on) revealed by their predecessors in the quest for unification. Nobody can know what richness may be revealed by the TOE, and the fragmentation it may engender.

And, please, no more citations of the defunct Superconducting Super-Collider: its physics will be covered ably by the Large Hadron Collider (LHC) at a fraction of the cost. Indeed, the LHC, at \$2 billion, is not so much more costly than the Spallation Neutron Source in the United States, at \$1.3 billion, or the Japanese Hadron Facility, just to take two examples of current accelerator projects motivated in large part by condensed-matter physics. Where would these, and synchrotron light sources, be if would-be particle unifiers had not developed the necessary accelerator tools?

Finally, I agree with Ellis that "isolationism is a prerequisite for successful physics", and with the authors of the book that it is a challenge for physics to "resist the temptation to project the false self-image of

a religion able to reveal the ultimate truth". However, I disagree with your headline that "there is no ultimate truth in grand unification". There is some of the ultimate truth, though certainly not all.

John Ellis

Theoretical Physics Division, European Laboratory for Particle Physics, CERN, CH-1211 Geneva 23, Switzerland

ECT damage is easy to find if you look for it

Sir—The reviewer of Max Fink's *Electroshock: Restoring the Mind*¹ claims that electroconvulsive therapy (ECT) "has proved to be one of the safest procedures in medicine" and that there is a "myth, largely promoted by anti-psychiatrists, that ECT damages ... brain functioning".

One can be sympathetic to psychiatry (as I am) and still imagine that passing 150 V between the temples to evoke a grand mal seizure might cause brain damage, especially when you realize that this 'cure' for depression requires this procedure to be repeated 10–20 times over a week or so. And when you talk to a friend who has been so treated and discover that a year later she is still experiencing huge gaps in recall of major life events, you begin to worry. Finally you discover that ECT's benefit is only temporary, so that many psychiatrists administer it chronically. Hmm.

Turn to the design of ECT protocols and you discover that many practitioners now administer ECT only unilaterally to the 'non-dominant' — non-verbal — hemisphere. Why? To avoid damaging the verbal hemisphere. In short, although ECT is completely safe, it is even safer when applied to the non-verbal hemisphere. Of course, equal damage is done to the non-verbal hemisphere, but it tells no tales.

ECT is used as an experimental tool by neuroscientists, as it releases massive quantities of glutamate, whose release following stroke causes significant neuronal death. Indeed, observers describe people who have had many ECT treatments as "punch drunk" — resembling boxers who have sustained chronic brain damage.

One reason psychiatrists are unaware that ECT is causing memory loss is that they do not test for it. Memory loss could be monitored by questioning patients before ECT about early events in their lives and then re-questioning them following each series of ECT. When this was done 50 years ago², memory losses were marked and prolonged. However, no effort has been made since to routinely perform this simple test.

It is a good bet that history will view ECT as one of what neuroscientist and author Elliot S. Valenstein calls the "great

and desperate cures" — and its promoters as kin to the promoters of lobotomy.

Peter Sterling

Department of Neuroscience, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6058, USA

1. *Nature* **401**, 327 (1999).

2. Janis, I. *J. Nerv. Mental Disease* **111**, 383–397 (1950).

Full effects of oil rigs on corals are not yet known

Sir—Bell and Smith in their Brief Communication¹ said that the azooxanthellate scleractinian *Lophelia pertusa* (L.) had been found on oil platforms in the North Sea. They also state that corals such as those on Brent Spar and those near the Beryl Alpha platform² "have been exposed to agreed quality standards of operational discharges, such as oily water, drilling muds and chemicals, and contaminants that may leak from cuttings piles...", suggesting that *L. pertusa* is thus not obviously affected by discharges from oil platforms.

But there is no definitive evidence that these corals have been exposed to any discharges, let alone a specific level of any "quality standard". It is perfectly feasible that their position in the water column precludes such exposure. Current understanding of the environmental sensitivity of deep-water species such as *L. pertusa* is limited by the lack of information on their biology and ecology. Our understanding of coral distribution is incomplete³; we know nothing of the reproduction or dispersal of these organisms, or their sensitivity to suspended sediments, or the effect on them of exposure to drilling discharges.

Bell and Smith's conclusion — that by leaving the 'footings' of large platforms with jacket weights of more than 10,000 tonnes in place, existing colonies will be preserved and *L. pertusa* might spread in the North Sea — is naive.

In the long term, we do not understand the ecological implications of leaving such structures in place, let alone whether they will survive to form some form of artificial reef. It is known that *L. pertusa* can settle on to man-made structures. For example, early linear extension rate measurements (*sensu* growth rate as described by Bell and Smith) were estimated from the overall length of corals that had settled on undersea cables⁴.

But we are a long way from understanding how any such decommissioning strategy would affect the ecology of the North Sea. An open debate and more effort to understand the underlying science are needed before statements can be made on the environmental sensitivity of any species.

J. M. Roberts

Scottish Association for Marine Science, PO Box 3,

Oban, Argyll PA34 4AD, UK

1. Bell, N. & Smith, J. *Nature* **402**, 601 (1999).

2. Pearce, F. *New Sci.* **164** (2211), 16 (1999).

3. Long, D., Roberts, J. M. & Gillespie, E. J. *Brit. Geol. Surv. Tech. Rep. WB/99/24C* (1999).

4. Wilson, J. B. *J. Mar. Biol. Assoc. UK* **59**, 165–177 (1979).

Cover adds fuel to the fire in evolution battle

Sir—We would very much appreciate an explanation by the editor as to why a part (the hands of God and Adam) of Michelangelo's painting *The Creation of Adam* was chosen for the cover of the 2 December 1999 issue of *Nature*. It was presumably meant to commemorate the elucidation of the first complete human chromosome nucleotide sequence.

However, the use of such Christian religious symbols to mark this event seems difficult to fathom. Does this harken back to the centuries old practice of natural theology¹? Does the elucidation of the human nucleotide sequence provide us with insights into the work of the Christian God at the creation event? Why not also use the Garden of Eden as the first event in the chronology of events leading to the revelation of the chromosome sequence in Fig. 1 of the News and Views article²? We are confident that the editors are well aware of the oft quoted statement by Dobzhansky: "Nothing in biology makes sense except in the light of evolution."

The decision to use this cover makes no sense. It is common knowledge that it continues to be a struggle in many parts of the United States to teach the principles of evolution in high schools. The recent decision by the Kansas Board of Education reveals the problem is not going to go away soon. Many letters to *Nature*, *Science* and other publications reveal the dismay felt within the scientific community regarding this action in Kansas. We are also dismayed by efforts in the United States to weaken the teaching of high-school science, but are also troubled by *Nature's* willingness to add fuel to the fire.

Gary C. Harris, Martina Königer

Biology Department, Wellesley College, Wellesley, Massachusetts 02481, USA

1. Mayr, E. *One Long Argument Charles Darwin and the Genesis of Modern Evolutionary Thought* (Harvard Univ. Press, Cambridge, Massachusetts, 1991).

2. Little, P. *Nature* **402**, 467–468 (1999).

Nature replies—The editorial staff debated this issue before publication and decided, with the enthusiastic agreement of the lead authors, that the image as a whole combined iconic symbolism with the science without implying that the Bible is true or that evolution is not the key to making sense of biology. — Editor, *Nature*

Seeking the great transition

Environmentally sustainable economies are unachievable without enhanced participation of the private sector. Scientists must facilitate this process.

**Gretchen C. Daily and
Brian H. Walker**

There is an urgent need, and a strong basis, for partnership between businesses and environmental scientists. Although this suggestion is not new — indeed, many people have pioneered visions for achieving sustainable development based on such a partnership^{1–6} — key players on both sides remain either opposed or remote.

Those who believe that ecologically destructive policies are a necessary, if unfortunate, outcome of achieving improved standards of living are wrong. This argument is analogous to the view held by some nineteenth-century businessmen that slavery was an essential element in the economic system⁷. Conversely, those who believe that business is not to be trusted, and should be uniformly opposed, condemn society to a dangerous and increasingly bitter conflict, the outcome of which is likely to be the worse for both the environment and society⁸. The transition to sustainability requires active cooperation between business and environmental science.

With the increase in globalization and

multinational corporations, the roles and capacities of governments in influencing the management of natural resources have diminished. They have declined to the level where most countries are unlikely to achieve the sustainable use of their resources without the active involvement of private enterprise. This is particularly evident in managing the Earth's natural capital so as to yield a flow of life-supporting services⁹. We agree with James *et al.*¹⁰, who recently argued that the true social costs of conservation are small: indeed, the benefits often outweigh the costs when the value of natural capital is properly calculated¹¹. But we do not concur that a small change in government policies or expenditures will do the job.

Rather, strengthening national policies on the environment may well achieve nothing — on its own — given the very limited ability to enforce such regulations in many regions of the world. For instance, even well-managed nature reserves are prone to rapid loss of species. Most governments lack the capacity to control the impact of human activity either within reserves (for example, poaching and its cascading effects) or outside their borders (affecting water, fire and

climate regimes, the influx of farm chemicals or invasive species, and so on)¹². The impacts involve diverse contributors across all scales, from local to global, and the failure to control one impact could hamper the successful mitigation of others. Such circumstances typify the environmental problems of today and demand a broad, integrated approach. Indeed, there is a growing trend towards private-sector involvement in the funding and management of nature reserves¹³.

What business can offer

Achieving environmental sustainability requires the insight and participation of business leaders, for many reasons. First, private enterprise is powerful, and has an enormous capacity to achieve sustainable use of resources. It is industry, not lawmakers, that manages and uses environmental resources, whether directly through production and consumption, or indirectly through influence on public opinion, governments, employment opportunities and consumer choices. Changing practices in industry — whether in the production of cattle, cotton, cars, entertainment or fashion — can produce cascading effects through all the determinants of environmental impacts.

Second, private enterprise is innovative and adaptable. The biological evolutionary process of reproduction, mutation/recombination and selection is mimicked far more quickly by cultural evolution in business than by any other form of human social organization. Fitness is measured as profit, and the corresponding evolutionary process involves production efficiency, innovation and selection (largely through competition). The challenge for society is to alter selection pressures so as to align profit-seeking with sustainability. This is an important role for government, one rarely played well. There is, however, also tremendous opportunity for self-interested action on the part of business in seeking that elusive alignment (see Box 1).

Third, private enterprise is efficient, at least relative to government. Its efficiency is geared towards minimizing private costs and thereby often imposes costs on society. In principle, however, this efficiency could be exploited to improve environmental conditions. Electrolux, the world's largest manufacturer of household appliances (it has annual revenues of US\$14.5 billion), estimates that 80–90% of its environmental impact occurs during the use of its products. The company's aim is thus to lead in materials and energy efficiency, and to

Box 1: Aligning market forces with the environment

The growing trend towards seeking competitive advantage through superior environmental performance is being supported by the emergence of consulting companies specializing in this area. Natural Strategies, Inc. (NSI) is one leading California-based company. NSI develops “strategic approaches to decision-making that use sustainability as a key driver”, according to co-founder Adam Davis.

For many corporations, sustainability is defined not only in terms of the natural environment, but incorporates a wide array of social and economic considerations²². NSI has developed the sustainability management system (SMS), which integrates issues of environment, health and safety, and social equity. The latter encompasses fairness in labour

practices, technology transfer, education of local communities, involvement of stakeholders and sensitivity to local culture. SMS combines broad conceptual frameworks for understanding sustainability, such as the ‘natural step’ (see Box 2, overleaf), with concrete and practical management tools.

One NSI client, Genencor International, Inc., is using this approach to develop comprehensive SMSs in nine facilities worldwide. These go far beyond what is required for complying with government regulations. Charles McGlashan, principal of NSI, describes a “growing willingness to adopt sustainability programs, especially when ideas are designed for implementation and a realistic return on investment”.

NSI has also been

instrumental in the development of the selection criteria used to evaluate companies for a new sustainability-oriented mutual fund called Portfolio 21. Developed by Progressive Investment Management, the fund goes beyond conventional approaches to socially responsible investment by translating sustainability frameworks into a detailed set of investment criteria and a comprehensive scoring system. Growing demand for thoughtful metrics in sustainability performance is driving improved analysis and screening of companies seeking reputational rewards. Efforts to promote sustainable behaviour in the marketplace have moved beyond vague notions and emotions into the realm of real work, performance and returns.

create consumer demand for such efficiency.

Finally, private enterprise is pragmatic. Business leaders can offer insight into the underlying socio-economic drivers of environmental change, the politically accessible leverage points, and actions that might lead to eventual solutions. The active participation of a critical mass of industries creating—and absorbing—particular environmental impacts could catalyse the implementation of effective countermeasures.

Can such a critical mass be generated? Efforts to address climate change serve as a useful model⁶. Twenty-one corporations (largely Fortune 500 companies, with US\$550 billion in collective annual revenues) have joined the Pew Center on Global Climate Change to research alternative courses of action. The group states: “We accept the views of most scientists that enough is known about the science and environmental impacts of climate change for us to take actions to address its consequences”¹⁴. Member corporations are also taking voluntary steps to reduce emissions now. DuPont, for instance, has reduced greenhouse-gas emissions from its global operations by 45%, improved energy efficiency by 15% below 1990 levels, and set ambitious targets for the future. Meanwhile, major corporations (first BP and Shell, now both working with the Pew Center, followed by Dow and Ford) have resigned from the Global Climate Coalition, a Washington-based group opposed to the Kyoto Protocol. The balance is clearly shifting towards co-operative engagement and early action.

The Sierra Business Council offers a region-based example of developing critical mass. Founded in 1994, it is an association of more than 550 businesses “working to secure the economic and environmental health of the Sierra Nevada”¹⁵. The region is the third most rapidly growing area in California, the world’s seventh-largest economy. To maintain the qualities that attract people and financial capital to it and which are under threat, the influential council is both developing innovative systems for measuring quality of life (including natural and social capital) and building a consensus around agendas to sustain these assets.

The role of environmental science

Many businesses are discovering profitable ways of reducing environmental impacts within the existing economic system. Bigger steps—changing the system to align profit-seeking with sustainability—require at least four ingredients: a scientifically based vision of a ‘sustainable’ business enterprise; an understanding of the competitive advantage in achieving the vision; effective communication of the vision; and a means of incremental implementation. The first of these requirements is where science can contribute most, in at least four ways.

Environmental science can provide

Box 2: Standards for the future

A suite of frameworks and tools have emerged in the past decade—most in just the past few years—to start making ‘sustainability’ operational in the corporate environment. The Natural Step programme, an overarching framework developed in 1989 in Sweden, is pre-eminent in Europe and now attracts major US corporations. It incorporates analytical tools from ecological footprint analysis²³, the ecosystem services framework⁹ and industrial ecology²⁴.

There is also a growing number of environmental (and related) standards, basically compliance procedures for guaranteeing to the clients of a business that a particular set of practices is actually being implemented. Certification is largely voluntary, driven by the marketplace. Individual countries or industrial groups often adopt such standards, however, making them de facto requirements. Examples include:

- **Eco-Management and Audit Scheme (EMAS):** a voluntary scheme introduced in 1995 to the European Commission by its member states. To qualify for certification, an organization must meet a set of verifiable standards and make certain information public.

- **International Standards Organization (ISO) 14001:** the pre-eminent US standard for environmental management systems, introduced in 1996; comparable to EMAS. The 1992 Rio conference was the impetus for developing these standards²⁵.

- **Social Accountability (SA) 8000:** introduced in 1998; to qualify for certification, an organization must meet verifiable standards of child and forced labour, health and safety, freedom of association, discrimination, disciplinary practices, working hours, compensation, and management systems.

- **Global Reporting Initiative (GRI):** this scheme is under development, and incorporates a wide array of environmental metrics based on annual financial reporting.

The emergence of these standards is a key development. There is, however, surprisingly little interaction between professionals working to reduce environmental impacts and those working to assess the status of the environment²⁴. The two groups have independently developed metrics to assess environmental performance, on the one hand, and ecosystem conditions, on the other. The lack of communication is a major barrier to progress, especially on more subtle and complex environmental impacts, such as genetic-modification technology or the status of ecosystem services. Interactive, long-term engagement of these groups is necessary to ensure that these standards are appropriate and effective.

predictive power. No one has a perfect crystal ball, but scientists are able to discern important trends. Many environmental changes that take society by surprise are actually foreseeable—only the market does not reliably transmit the signals they send. Declines in the capacity of the atmosphere or a water body to absorb wastes can proceed largely unnoticed for years until a crisis erupts (for example, the Antarctic ‘ozone hole’, the Gulf of Mexico ‘dead zone’). There is great commercial value in knowing about these changes before the market or the media get around to announcing them, or governments pass regulations on them.

Science can also assist in anticipating, and withstanding the shock of, genuine surprises. The science community typically learns of new issues years, sometimes decades, before they appear on the radar screens of governments, non-governmental organizations and the media (examples include the loss of antibiotic potency, climate change, nitrogen saturation and valuing ecosystem capital). The scientific process strongly favours conservatism in the reporting of information. Paying attention to the scientific radar screen is key to creating, and maintaining clarity in, a corporation’s long-term vision.

Second, environmental science can facilitate adaptation. When new issues do emerge

publicly, much more flexible, efficient and innovative ways of addressing them can be devised with joint business/science communication, and are likely to appear in the form of government regulation, than in its absence. Businesses can profit by staying ahead of the regulators, reducing environmental impacts voluntarily⁵. Many new approaches are facilitating this smart manoeuvring (see Box 2).

There remains a long way to go, especially in the spatially extensive primary industries such as agriculture, forestry, pastoralism and mining. But even here there are promising signs. One example is the Collins companies, which produce wood products including lumber, veneer, plywood and particleboard. Collins was the first timber company in the United States to operate under the rules set by the Forest Stewardship Council (begun in 1993). Collins has also implemented training for its employees in the principles of sustainability. This has resulted in reductions in the industry’s impact on the environment through measures such as better procurement of raw materials, improved energy efficiency, waste reduction and water conservation.

Third, environmental science can facilitate prioritization. There are myriad ways of protecting the environment, from recycling office paper to entirely recreating a

corporation's vision (in the case of BP, for example, from an 'oil company' to an 'energy services company'). The actions that are most effective will be determined at the level of the individual corporation and below. Environmental science can help to quantify the relative merits of alternative courses of action, to identify critical leverage points and opportunities for intervention, and to develop the full scope of options.

Finally, increasing numbers of environmental scientists are becoming involved, and are acquiring the skills to convey and apply their knowledge in a practical way. Much of the calculation for business involves the transaction cost — the value of the information itself weighed against the time, effort and expense required to get it. This is dawning on the environmental science community, as reflected by the creation in 1999 of the Aldo Leopold Leadership Program. This US-based programme trains academic scientists to communicate policy-relevant information more effectively to the media, industry, organizations and government. It can accommodate 20 scientists a year, and is vastly oversubscribed.

The time is ripe

There are many signs of opportunity, paramount among them the fact that a growing number of business leaders are acutely aware of and genuinely concerned about the state of the environment. Some have taken a strong public stand, revealing diverse motivations ranging from pure pragmatism (as in Scandic Hotels' early (1994) recognition of the market edge associated with environmental sustainability) to profound philosophical transformation (as in the case of Ray Anderson, chief executive of the Interface carpet company, also in 1994). Many others are working deliberately behind the scenes; although their efforts are not publicly apparent, they involve serious commitments.

Other signs include: the increasingly positive responses of shareholders and the financial community (for example, the position of 'green' trusts among top trust companies); the (belated) recognition by industry of consumer reaction to the aggressive application of technology related to genetically modified organisms; the growth in interdisciplinary collaboration, especially between ecologists and economists; the increasing acceptability in academic institutions of 'outreach' beyond the ivory tower to government, the private sector and the public; and, simply, the growing number of companies involved (for example, the 125 or so companies — most of them multibillion-dollar transnational corporations — that are engaged in the 'eco-efficiency' programme of the World Business Council on Sustainable Development).

Is it naive to foster this partnership? Will scientists be exploited for the value they

add to a false green corporate image? Will corporate leaders turn away because of the difficulty of converting scientific ideas into the hard recommendations and decision rules that businesses need to operate? Will either side pay sufficient attention to social (equity) dimensions of environmental sustainability^{16–18}? Is there sufficient trust in any quarter to move forward?

Admittedly, there is a long way to go. Especially where the initial costs are high, few corporations will act unilaterally. In the cut-throat business of attracting and satisfying shareholder investment, reducing short-term profits is only possible on a level playing-field of environmental requirements. There will always be short-sighted and corrupt companies and governments; if unrestrained, their actions will prevent the necessary change in overall corporate behaviour.

The challenge is to tighten feedback loops so that the environmental consequences of individual or corporate behaviour "quickly come home to roost"¹⁹. But the absence of appropriate feedbacks is not easy to remedy. Consider the prevalence of environmentally destructive subsidies, which often benefit the private sector precisely by preventing market feedbacks from operating²⁰. Such subsidies have been devastating in the fishing industry. Virtually every major fishery in the world has been drastically degraded, some irreparably. If complete scientific knowledge had been available to the industry at the outset, would it have regulated itself to prevent the over-fishing that took place? Not without strongly enforced legislation. Major corporations committed to long-term involvement support such legislation.

Clearly, government, environmental and other organizations, the media and the general public have major roles in establishing appropriate feedback mechanisms. Indeed, they deserve much credit for the success achieved so far. A concerted effort by strong leaders, with the necessary institutional support, is required for a critical mass of corporations to adopt sustainability as integral to their mission. Thereafter, the non-sustainable behaviour of the remaining corporations will begin to be regarded as aberrant, and as jeopardizing the future profitability of the industry.

Proposal

Private enterprise is a system of which we are all willing beneficiaries. Imagine running a thread from each ingredient of one's daily, material consumption to its point(s) of origin in the natural environment: the planet would be criss-crossed with threads even before our breakfast cereals had been accounted for, and most of these threads would meander through myriad corporations. Basically, as long as it is profitable to engage in environmental destruction, collectively we will do it. Indeed, refocusing

consumer desires may be the biggest challenge of all²¹.

Radical environmentalists will accuse us of dancing with the devil. But we assert, rather, that the time is ripe for a synergistic partnership to forge the great transition to a sustainable future. Involvement of corporations certainly doesn't guarantee anything. But failing to engage industry may well ensure that sustainability remains largely an academic debate. ■

Gretchen C. Daily is at the Center for Conservation Biology, Department of Biological Sciences, Stanford University, Stanford, California 94305-5020, USA.

Brian H. Walker is at CSIRO Wildlife and Ecology, GPO Box 284, Canberra, ACT 2601, Australia.

Acknowledgements We are grateful for comments from Scott Daily, Paul Ehrlich, Larry Goulder, Paul Hawken, Donald Kennedy, Jeff McNeely, Harold Mooney, Peter Raven, Virginia Walbot and Gideon Yoffe. The idea for this article arose at a meeting of the Beijer International Institute for Ecological Economics.

1. Schmidheiny, S. *Changing Course: A Global Business Perspective on Development and the Environment* (MIT Press, Cambridge, Massachusetts, 1992).
2. Hawken, P. *The Ecology of Commerce: A Declaration of Sustainability* (HarperBusiness, New York, 1993).
3. Socolow, R., Andrews, C., Berkhout, F. & Thomas, V. (eds) *Industrial Ecology and Global Change* (Cambridge Univ. Press, 1994).
4. Richards, D. J. (ed.) *The Industrial Green Game: Implications for Environmental Design and Management* (National Academy Press, Washington DC, 1997).
5. Hawken, P., Lovins, A. & Lovins, L. H. *Natural Capitalism: Creating the Next Industrial Revolution* (Little, Brown, New York, 1999).
6. Romm, J. J. *Cool Companies: How the Best Businesses Boost Profits and Productivity by Cutting Greenhouse Gas Emissions* (Island Press, Washington DC, 1999).
7. Dunphy, D. & Griffiths, A. *The Sustainable Corporation: Organizational Renewal in Australia* (Allen & Unwin, Sydney, 1998).
8. Wilkinson, P. *Nature* **396**, 511–512 (1998).
9. Daily, G. C. (ed.) *Nature's Services: Societal Dependence on Natural Ecosystems* (Island Press, Washington DC, 1997).
10. James, A. N., Gaston, K. J. & Balmford, A. *Nature* **401**, 323–324 (1999).
11. Daily, G. C. *et al.* *Science* (submitted).
12. Hackel, J. D. *Conserv. Biol.* **13**, 726–734 (1999).
13. McNeely, J. A. *Mobilizing Broader Support for Asia's Biodiversity: How Civil Society Can Contribute to Protected Area Management* (Asian Development Bank, Manila, 1999).
14. *Joint Statement of the Business Environmental Leadership Council* <http://www.pewclimate.org>
15. Sierra Business Council *Sierra Nevada Wealth Index: Understanding and Tracking Our Region's Wealth* (Sierra Business Council, Sacramento, California, 1996).
16. Raven, P. H. in *Frontiers of the Mind in the 21st Century* (ed. Gifford, P.) (in the press).
17. Daily, G. C. & Ehrlich, P. R. *Ecol. Appl.* **6**, 991–1001 (1996).
18. National Research Council *Our Common Journey: A Transition Toward Sustainability* (National Academy Press, Washington DC, 1999).
19. Levin, S. A. *Fragile Dominion: Complexity and the Commons* (Perseus, Reading, Massachusetts, 1999).
20. Myers, N. & Kent, J. *Perverse Subsidies: Tax Dollars Undercutting Our Economies and Environments Alike* (International Institute for Sustainable Development, Winnipeg, 1998).
21. Rosenblatt, R. (ed.) *Consuming Desires: Consumption, Culture, and the Pursuit of Happiness* (Island Press, Washington DC, 1999).
22. Elkington, J. *People Planet* **8**, 10–11 (1999).
23. Wackernagel, M. & Rees, W. *Our Ecological Footprint: Reducing Human Impact on the Earth* (New Society Publishers, Gabriola Island, British Columbia, 1996).
24. Schulze, P. C. & Frosch, R. A. in *Measures of Environmental Performance and Ecosystem Condition* (ed. Schulze, P. C.) 1–12 (National Academy Press, Washington DC, 1999).
25. Sasseville, D. R., Wilson, W. G. & Lawson, R. W. *ISO 14000 Answer Book: Environmental Management for the World Market* (Wiley, New York, 1997).

Is the sky made from pi?

Our concept of the Universe is prey to various numerical interpretations.

Just Six Numbers: The Deep Forces that Shape the Universe

by Martin Rees

Weidenfeld & Nicolson: 1999. 173 pp.
£12.99, \$22

The Nine Numbers of the Universe

by Michael Rowan-Robinson

Oxford University Press: 1999. 188 pp.
£16.99, \$32.50

Frank Wilczek

More than two millennia have passed since Pythagoras proclaimed, "All things are Number". Was his proclamation reverie or revelation? In the books reviewed here, two distinguished cosmologists, from profoundly different perspectives, weave popular accounts of their subject around this question.

Developments in physics and astronomy over the past century have brought the Pythagorean vision into sharp focus. The homogeneity of the Universe, and the sameness (or 'universality') of the physical laws throughout, are established by observation. Also, we have a well-tested picture of evolution from a hot, dense, extremely homogeneous phase — the Big Bang. We can codify the physical laws in a remarkably simple, beautiful and mathematically precise set of equations — the so-called Standard Model, plus general relativity.

To fully specify cosmology requires at least the amplitude and slope of the spectrum of primeval fluctuations, the mass fraction in one or several kinds of dark matter, the baryon fraction, the Hubble parameter and the value of the cosmological term. Similarly, to fully specify the Standard Model (including all the heavy quarks and leptons) requires many logically independent parameters.

Thus, our fundamental working models of nature, although containing an elegant and powerful Pythagorean core, begin to look complicated and unsatisfying when fully fleshed out. Moreover, we have learned that there are many questions we should not expect to be able to answer from first principles. Whereas it was reasonable for Johannes Kepler to try to fit the ratios of planetary orbits using ideal mathematical constructions (regular polyhedra), we now understand that these ratios are highly contingent facts, accidents of a unique history.

The title and introduction of Martin Rees's book led me to expect something quite different from what the rest of the book delivers. I had expected to find a celebration of the triumph of Pythagorism — an

account of how, given the values of just a few numbers (say 6), you can account in detail for a wealth of physical phenomena and, on the way, do some pretty impressive engineering, by pure calculation. Instead, Rees delivers six case studies of how the Pythagorean programme very nearly goes catastrophically wrong. If any of six specific, chosen quantities had been significantly different, Rees argues, our world would be unrecognizable and intelligent life would most probably be impossible.

Two examples will give a flavour of the argument. The binding energy of protons and neutrons into helium-4 is 0.007 of their rest mass. This number arises, according to modern elementary particle physics, from a rather intricate interplay between the masses of 'up' and 'down' quarks (relative to the basic quantum chromodynamic scale) and the fine-structure constant. From the point of view of fundamental theory it is a remote epiphenomenon, and without absurdity one can easily imagine worlds in which it had a slightly different value. Yet, Rees argues, if it were as small as 0.006, deuterium would be unbound and no elements other than hydrogen would be created; while if it were as large as 0.008, stellar evolution would be so fast and violent that life as we know it could not evolve.

The ratio of the strength of gravitational to electric forces, for protons, is about 10^{-36} . This extraordinary number arises, according to a rather more uncertain extrapolation of modern elementary particle physics, as a result of the slow running of the strong coupling constant. If it were significantly larger, gravity on bound systems such as planets would be much more oppressive, and large multicellular creatures would be crushed by their own weight (also stars would evolve faster). If it were significantly smaller, planets would not form in the first place.

In documenting these examples, Rees is moving towards a deeply subversive position. It is becoming increasingly difficult to believe, he argues, that the complex 'fine tunings' of physical parameters that seem to be necessary for the emergence of intelligent life are unique consequences of a simple fundamental theory. He favours instead the idea that we live not in a Universe but rather in a Multiverse, which contains regions with drastically different properties (for example,

Taking a count: Pythagoras believed that "All things are Number".

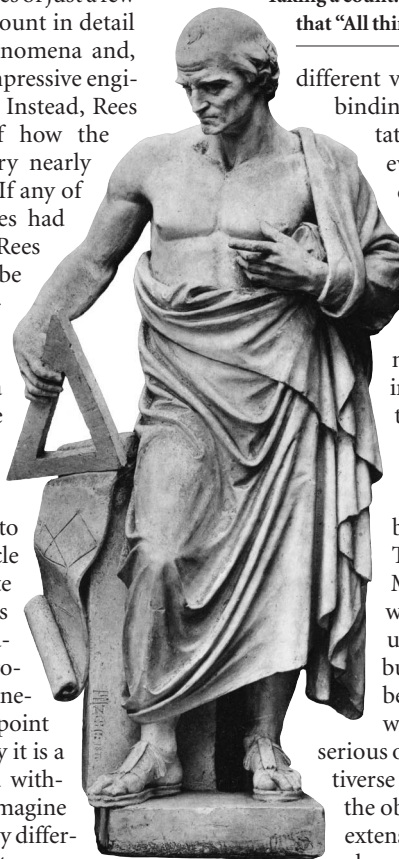
different values of the nuclear binding fraction, gravitational strength, or even space-time dimensionality). In this framework, the 'fine tunings' could be explained anthropically. They need not have occurred, indeed in most places they do not occur, but where they don't there's no one around to watch the botch!

Twenty years ago the Multiverse concept would have seemed utterly far-fetched, but now it threatens to become conventional wisdom. The most serious objection to the Multiverse is, of course, simply the observational fact that extensive astronomical observations of distant

regions of the cosmos have disclosed the basic uniformity, not diversity, of physical law. But the theory of cosmic inflation has made it plausible that the portion of the world we can currently observe might be just a small portion of the whole, an initially tiny (and hence uniform) patch inflated to gigantic proportions. And high-energy theorists working on supersymmetric unified theories and string/M theory find themselves confronted with a bewildering variety of apparently consistent solutions, with nothing to choose between them. Might each have its homeland?

If these ideas are correct, then the irreducible element of contingency we noted above extends much further than commonly allowed. Several of the seemingly 'fundamental' parameters of physics would, in fact, be features of our environment. They could never be calculated directly from fundamental theory without taking a serious detour through anthropism, because they would be consequences of our position in the Universe amidst the larger Multiverse, not of any universal truth.

Thus, Rees is the anti-Pythagoras. Whether or not they are convinced by his major thesis, readers will find Rees's short,



BETTMANN/CORBIS

book reviews

well-written book an enjoyable and provocative intellectual adventure.

Michael Rowan-Robinson's book represents, by comparison, down-to-earth cosmology. He discusses the key observations — past, present and future — in much greater depth, including frequent and strenuous warnings about their uncertainties, yet he does so concisely. In the penultimate chapter he discusses his own involvement in investigating the starburst-galaxy phenomenon. Although perhaps not quite as grand as the other material, it is in compensation fresh and personal. The two books are in many respects complementary and, by looking at both, a reader could get an excellent, rounded view of the exciting state of contemporary physical cosmology.

Neither of these books, with their heavy focus on the wild-and-woolly frontiers of cosmology, begins to do justice to the truly remarkable triumphs of Pythagorism closer to home. Given five pure numbers — the electron, up- and down-quark masses in units of the quantum chromodynamic scale, plus the fine-structure constant — one can accurately account for all the phenomena of chemistry, and the structure of ordinary matter. Add a couple more — Newton's gravitational and Fermi's weak-coupling constants — and essentially all of astrophysics and most of cosmology enter the charmed circle of understanding. Small parts of this

great scientific success story have been told, but it has yet to find its Milton. ■

Frank Wilczek is at the Institute for Advanced Study, School of Natural Sciences, Olden Lane, Princeton, New Jersey 08540, USA.

Arctic antics

Ice Finders: How a Poet, a Professor, and a Politician Discovered the Ice Age

by Edmund Blair Bolles

Counterpoint: 1999, 256 pp. £16.50, \$24

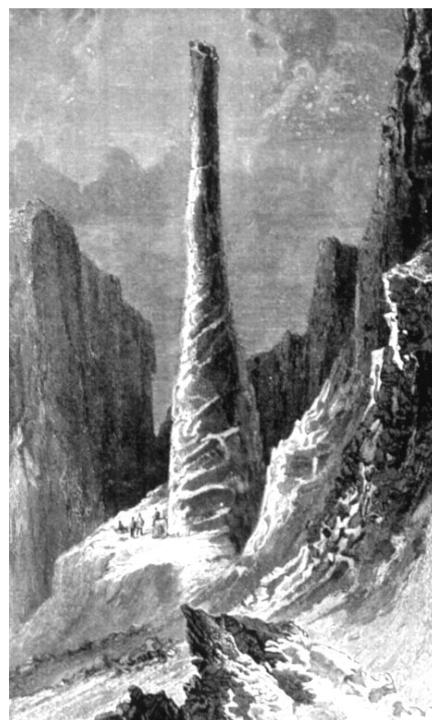
Douglas Palmer

The Ice Finders by Edmund Blair Bolles is about the struggle for scientific acceptance of the evidence for a recent Ice Age. Although a good idea for a popular science book, it is curiously constrained by its alliterative subtitle — the 'poet' is Elisha Kent Kane, the 'professor' Louis Agassiz and the 'politician' is Charles Lyell. Kane was, in fact, an assistant surgeon in the US Navy, who led the 1853–55 US expedition of the *Advance* in search of Sir John Franklin and the existence of an open, ice-free Arctic Ocean. Agassiz, the Swiss expert on fossil fish and glaciation, was indeed a professor at the Academy of Neuchâtel and subsequently Harvard, and the British 'encyclopaedic' geologist Lyell was one of the best-known Earth scientists of the time.

Bolles sets out his methodological 'stall', in a chapter entitled "Ignorant, ambitious men", by declaring that "the most famous law of mechanical intelligence says: Garbage in, garbage out". Consequently, Lyell's book (*The Principles of Geology*, 1830) becomes, "as far as the Ice Age was concerned ... another case of garbage in". To be fair, Bolles is writing popular science, but I find such simplistic clichés irksome. I doubt whether they are really necessary to 'hook' the reader if there is a good story to tell, and basically there is a good story here.

The conceit of the plot lies in the interweaving of the narrative of Kane's expedition with the more complicated stories of the discovery and interpretation of glacial phenomena by Agassiz, Lyell and their contemporaries at least a couple of decades earlier. Kane's story is well told, but has the feel of being shoehorned into the narrative. Presumably Bolles, an American, wanted a 'homegrown' hero as a foil to the predominance of European characters. Eventually, the strands are pulled together when Kane returns and publishes an account of his travels and scientific observations (*Arctic Explorations in the Years 1853, '54, '55, 1856*), with an encomium by Agassiz.

In order to make the Kane story more relevant to the plot, Bolles has to downplay earlier Arctic exploration and public awareness of polar ice fields. He claims that "most

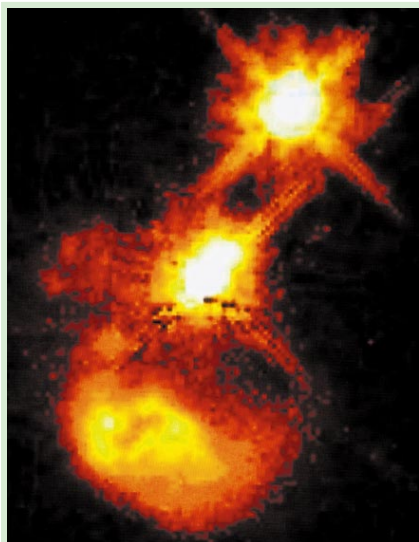


Polar prominence: Kane's sketch of a column of greenstone encountered during his explorations.

people never saw, never heard of the great ice. Presumably Erik the Red and a few other Vikings who made it to Greenland had some idea." However, Clive Holland (*Arctic Exploration and Development c. 500 BC to 1915*, Garland, 1994) estimates that the Viking population of Greenland may have numbered as many as 3,000 around the first millennium. Holland also lists hundreds of expeditions to the Arctic before the 1820s by, among others, Russian, Scandinavian, French and British explorers.

As for public awareness, "The ice was here, the ice was there, The ice was all around; It cracked and growled, and roared and howled, ..." — Samuel Taylor Coleridge's *The Rime of the Ancient Mariner*, published in 1798, was all around as well. The poem enjoyed enormous success and subsequent influence. In addition, the *Quarterly Review* of 1817 and 1818 devoted two long articles to Arctic exploration which, along with Coleridge's poem, are thought to have stimulated Mary Shelley's use of the Arctic in her novel *Frankenstein, or, The Modern Prometheus*, published in 1818. Her explorer, Captain Robert Walton, writes that, "We are still surrounded by mountains of ice, still in imminent danger of being crushed in their conflict. The cold is excessive ... Frankenstein has daily declined in health ..." Surely, there is an interesting story here about the growing public fascination with the polar regions from the turn of the century.

The real meat of the Bolles story is the discovery and argument over the nature of glacial phenomena such as erratics, parallel roads and scratched rock surfaces. This part



Quasars, nebulae, and more

With the Hubble telescope quasars can be imaged in great detail, such as quasar IRAS 04505–2958, some 3 billion light years away, shown here. The image comes from *Deep Space: New Pictures From the Hubble Space Telescope* (Constable, £16.99), a collection of recent Hubble images, with an explanatory text by Simon Goodwin and John Gribbin.

of the history documents the difficulties most scientists experience in seeing phenomena as they actually are, unfettered by preconception. The description of the way in which Agassiz, Lyell and other interested scientists such as William Buckland, Roderick Murchison and James Forbes responded to new ideas and phenomena is intriguing. Agassiz was particularly egotistical; he only welcomed Forbes' support of his ideas until the latter's acute observation and questioning threatened his pre-eminence. Generally, Bolles' 'factionalization' seems convincing, although I am not sure how professional historians of these events will respond. Asides such as "thanks to Agassiz, they (Britain's geologists) had clear illustrations and facts to help them resolve the chronology of their local formations", which promotes Agassiz's identification of fish fossils way above their worth, leaves a certain sense of unease.

It is always difficult to find the right moment to end a historical narrative. Bolles chooses 1863, when Lyell finally accepted that continental ice sheets on a vast, Greenland-like scale had occurred in Europe and North America during the Ice Age (in *The Geological Evidences of the Antiquity of Man*). But in many ways the story was just beginning. In 1855, A. C. Ramsay claimed that some Permian-age conglomerates in England had been transported by glacial action, and the following year W. T. Blanford made the same claim for contemporaneous sediments in India — the Ice Age was not unique. And in Glasgow, a brilliant, young and largely self-taught Scot, James Croll, was sowing the intellectual seeds for an explanation of ice ages.

Douglas Palmer is a science writer, and is at 31 Mawson Road, Cambridge CB1 2DZ, UK.

The Osler magic revisited

William Osler: A Life in Medicine

by Michael Bliss

Oxford University Press: 1999. 600 pp.

£27.50, \$35

W. F. Bynum

Louis Pasteur was once described as the most perfect man ever to enter the kingdom of science. The kingdom of medicine award would undoubtedly go to Sir William Osler, certainly if English-speaking clinicians over the age of 50 constituted the electoral college. There may not be so many Osler societies, medals, lectureships, prizes and buildings as there are *rues* Pasteur, but the French have always been more susceptible to the Great Man syndrome than have the Anglo-Saxons. The wonder is that, during Osler's life and after, so much fuss was made about a man who did not discover

anything of much importance.

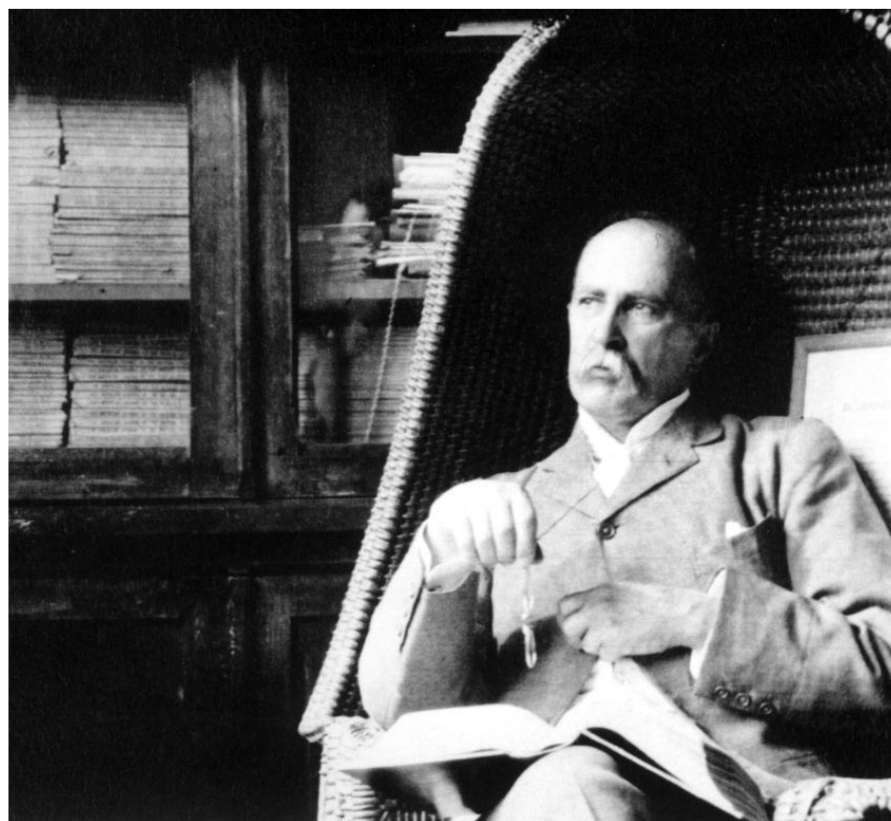
If the doctors have always adored him, historians have tended to disparage Sir William as emblematic of all that is wrong with the cosy, self-satisfied and self-serving world of élite medicine. He even seemed living proof of the cogency of his own quip that men do not achieve much of significance after the age of 40, and are all washed up by 60, when they ought to be forcibly retired, if not chloroformed. (He believed that women mature later.) His own seventh decade was spent as Regius Professor of Medicine at the University of Oxford, seeing a few private patients, sitting on committees, collecting books, delivering inspirational speeches, studying medical history, entertaining on a large scale, and making a modest medical contribution to the Allies' war effort. All very energetic and even admirable; whether it is the stuff of idolatry is not self-evident.

The vast literature on Osler has almost always been written by adoring doctors. Much of it is derivative on the massive *Life of William Osler* (1925), which won a Pulitzer prize for its author, the brilliant American neurosurgeon Harvey Cushing. Cushing had been Osler's junior colleague at Johns Hopkins University, where he had gained his surgical experience covering for his chief, William Halsted, when the latter was away operating on private patients or, more commonly, incapacitated through his severe cocaine addiction. Osler, but not Cushing, knew of Halsted's affliction, which he had

acquired experimenting with the use of cocaine as a local anaesthetic. Cushing adulated Osler and was even at Revere Osler's bedside when this only surviving offspring died of shrapnel wounds in the killing fields of the First World War.

Although Cushing's *Life* has always provided the gold standard, adoring fans have subsequently combed through almost every facet of Osler's career, personality and achievements. The bibliography of writings about Osler has gone through two editions. This vast output has been made more accessible by the Osler Library at McGill University, to which he left his magnificent collection of books and papers, and where, fittingly, his ashes rest. By contrast, Pasteur's remains lie in the research institute he created.

Michael Bliss's biography is one of the first major contributions to the Osler *oeuvre* by a professional historian. Surely he could be expected to do to Osler what the historians have done to Pasteur, Isaac Newton, Albert Einstein and numerous other scientific heroes — find the feet of clay. After all, Bliss had some sharp things to say in a previous biography about a fellow Canadian, Sir Frederick Banting (*Banting: A Biography*, University of Toronto Press, 1992). Those looking for scandal, however, will be disappointed. Bliss even scotches the rumour that Osler, who married late, might have had an affair with a cousin during his McGill days. Osler revisionism, Bliss informs us, simply does not work.



Saint William: considered by many to be the embodiment of humane, scientific medicine.

WELLCOME INSTITUTE LIBRARY

New in paperback

Made to Measure: New Materials for the 21st Century

by Philip Ball

Princeton University Press, \$17.95, £11.50

"Many books on new materials have more hyperbole than substance; this is not one of them. Ball writes of his materials with enthusiasm, but tempered realism. He recognizes that most (but far from all) new materials will be passing wonders." A. Lindsay Greer, *Nature* **392**, 561–562 (1998)

The Age of Spiritual Machines: When Computers Exceed Human Intelligence

by Ray Kurzweil

Penguin, \$14.95

"Kurzweil's account focuses mostly on the intelligence of machines and the way humans will interface with them via direct neural connections to our brains. The writing is always lively and the arguments are well presented and amplified by numerous boxed cut-outs, tables and graphs showing the ever-accelerating pace of computing technology ... [Recommended] to any reader who wants a mind-expanding account of the rise of the age of intelligent machines." John L. Casti, *Nature* **397**, 663–664 (1999)

Mapping the Mind

by Rita Carter

Seven Dials, £14.99

"Information there is aplenty, nugget after nugget of it to stimulate and tease the most jaded palate." Jeffrey Gray, *Nature* **399**, 652 (1999)

Instead, Bliss succumbs to the Osler magic, and offers us an immaculately researched and finely written account of this most famous of all modern doctors. It goes beyond Cushing's 'life and letters' approach, but Bliss reinforces rather than displaces the received image of Saint William. He sticks close to the rich manuscript sources and avoids psychologizing. We follow Osler through his career, from his childhood as the son of an Anglican parson in backwoods Ontario to education in Toronto and Montreal, where he graduated in medicine from McGill. Bliss remarks, almost in passing, that Osler was never comfortable with any language but English, despite his education in predominantly francophone Montreal, and his active participation in numerous international medical congresses at a time when French and German were more common languages of scientific currency.

A rich brother helped Osler spend a couple of postgraduate years in Europe, which was the making of him. From then on, it seemed almost effortless: chairs at McGill, the University of Pennsylvania, Johns Hopkins and Oxford, with offers of serious nibbles elsewhere, including Harvard,

Toronto and Edinburgh. Everyone wanted Osler, even before his *Principles and Practice of Medicine* (1892) made him the public embodiment of humane, scientific medicine. Not a particularly effective public speaker, he was nevertheless much in demand, and occasionally moved his audience to tears. He was at his best at the bedside, however, both as a teacher and as a doctor. He inspired confidence in his patients and simply inspired his students.

There is nonetheless the nagging question of why Osler's reputation did not simply quietly fade away, as his students and colleagues followed him to their graves. Bliss's last chapter describes, even if it does not entirely explain, the posthumous Osler. It may be something as vague as 'character', or it may be related to his very ordinariness, or his irrepressible sense of fun. Maybe it is because, like him, we all really enjoy the company of children more than that of their stuffy parents. Whatever it is, the Osler phenomenon is still alive and well, even if Sir William is not the only saint to be have attracted acolytes for less than saintly motives. ■

W. F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BE, UK.

Seeing the world in a mote of dust

Dust: A History of the Small and the Invisible

by Joseph A. Amato

University of California Press: 2000. 288 pp.

\$22.50, £13.95

Ralph Lewin

According to some religions, the human race was created from a handful of dust. According to others, perhaps more plausible, that's how we'll all end up. So the subject of dust should be of some interest to all of us, even if we are not cleaners, librarians, vacuum-cleaner sellers or members of Alfred Doolittle's fraternity of dustmen in the film *My Fair Lady*.

I was glad to have this book sent to me to review, since dust has accompanied me throughout my life. As a child in London I used to watch golden motes of cotton dust spinning slowly as they fell through all-too-infrequent sunbeams. I survived many dusty weeks in San José, Costa Rica, when Volcan Irazu erupted. Every time the wind blew in the wrong direction, it shed so much ash over the city that lawns became grey, and dust-covered coffee plants died for lack of adequate photosynthesis.

I expected to find plenty of scientific information on abundances, sizes, velocities and compositions of interstellar particles,

including even little diamonds; of the rates of fall of aerosol particles and their scattering of light in the atmosphere; of redder sunsets on smoggy days; of the aeolian transport of ferruginous dust from the Sahara across the Atlantic Ocean to the Caribbean; of lead dust along roadsides and how it affects traffic policemen and children in areas where automobile fuels are (or were) leaded; of pollen and mites' eggs and allergies; of asbestosis leading to lung cancer; of domestic dust, and how static charges cause it to settle sideways on walls over radiators. Alas, I found almost nothing along these lines.

Amato writes well; he is a littérateur, an elegant stylist and, presumably, a good historian. I learnt about the craze for gold dust, and of rural tragedies in the American dust bowl during the 1930s. Sprinkled unobtrusively through the text are hundreds of little numbers, so small that they don't interrupt one's flow of reading; they refer to original sources quoted in 40 pages near the end of the book, as is perhaps proper for historical theses. (Maybe we scientists could copy this system.)

Clearly, though, Amato is no scientist. What little science we find here is not of the highest order. "A grain of musk perfumes a room for years, and a single grain of indigo colors a ton of water," he writes. "Antbread is a barely visible part of a tiny seed that ants drag to their nests and which, if uneaten, springs up into plants. ... In grain elevators ... a single spark from any source (even the tiny amount of electricity given off by the human body) can trigger an immense explosion." And so on. Amato has spread his net widely. His concept of dust seems to embrace not only the fluff that I surreptitiously sweep under rugs but also sawdust, powdered herbs, flour, gunpowder, dirt, mud, germs of all sorts, worms, lice, and even halitosis! His topics include a scattershot on cancer, Dolly the ewe, Nazism, plumbing, poverty, transistors, Vincent van Gogh ... but wait! Underneath all these dusty matters there is, in fact, yet another interesting and informative history of art, culture — yes, even science — in Western Europe during the past millennium.

I am reminded of a game we played in our school debating society. We put into a hat many bits of paper bearing different words as topics, such as God, Macbeth, underpants, UFOs, dust, and so on. Then one of us drew out a couple at random, was allowed two minutes to consider them, and then three minutes to give a short talk somehow integrating the two selected topics. Amato, if he ever played it, would have excelled.

I'm sure many lay readers will find this book entertaining. The University of California Press must have thought so, too. ■

Ralph Lewin is at the Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0202, USA.

Tycho and the ton of gold

Why a brilliantly conceived research programme failed.

Owen Gingerich

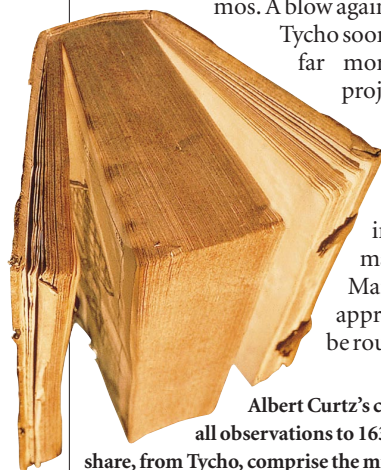
Tycho Brahe is widely known as the eccentric sixteenth-century Danish nobleman with the silver nose, the egotistic builder of a battery of pre-telescopic astronomical instruments, and the observer whose precious and refined data on Mars made possible Johannes Kepler's discovery of the elliptical orbits of the planets.

Tycho boasted that his observations on the small island of Hven had cost the King of Denmark more than a ton of gold. The gold bought an array of quadrants, sextants and armillary spheres, with which he could measure an accuracy of one minute of arc, essentially the limit of the naked eye. Tycho never published the great bulk of his observations, nor his rationale for making many of them.

In our world of publish or perish, an understanding of what drove Tycho has nearly perished, leaving only a caricature of one of the most brilliant experimentalists of the past millennium. Had the dimensions of the Solar System been as small as everyone from the ancient Greeks onwards had assumed, Tycho would be known as the towering observational cosmologist who was the first to demonstrate that Ptolemy's geocentric arrangement of the planets was untenable.

In 1572, a spectacular new star appeared in the constellation Cassiopeia, what would be recognized today as a supernova. Tycho promptly discovered that this star showed no diurnal parallax, the small angular displacement that depends on where from the Earth the observation is made. For the Moon, the parallactic angle is about a degree; since the nova's was much less than this, the new star had to be much farther away than the Moon. This placed it in the supposedly unchanging and incorruptible celestial regions of the cosmos. A blow against Aristotle!

Tycho soon decided on a far more ambitious project: to find the diurnal parallax of Mars. He believed that, in the Ptolemaic system, Mars at closest approach would be roughly 20 times



Albert Curtz's compilation of all observations to 1633. The lion's share, from Tycho, comprise the middle block.



The 1585 equatorial armillary, the most versatile and impressive instrument on Hven.

farther than the Moon. But if the Copernican system held, it would be only about ten times farther, making for an easy measurement of around 5 arcmin. Tycho set out to measure this parallax when Mars passed close to the Earth in the winter of 1582–83.

What Tycho did not know was that every astronomer since antiquity had underestimated the distance to the Sun by a factor of 20. This means that the parallax of Mars is actually around 20 arcsec, well below the resolution of the naked eye. Not surprisingly, Tycho at first found no evidence of the parallax. However, he did not give up, trying again in 1585. The results were disconcerting: Mars seemed to have a negative parallax, placing it beyond infinity, clearly an absurdity. Something was wrong, and Tycho soon identified the culprit. Atmospheric refraction had distorted the planet's position.

Meanwhile, Tycho improved his instruments. A subterranean observatory would protect his sextants and quadrants from the wind and allow for more stable mountings. By the 1587 opposition of Mars, the new Stjerneborg Observatory next to his Uraniborg castle was ready. Tycho's results were, by modern standards, astonishing. He claimed he had found the 5-arcmin parallax, and the Ptolemaic system had to be rejected. We know that Tycho couldn't have found a diurnal



The mural quadrant with scenes from Tycho's scientific establishment at Uraniborg castle.

parallax because the Solar System is much bigger than he supposed. Was he lying about a result he yearned so deeply to find?

The problem lay in a faulty table of atmospheric refraction. As a conscientious observer, Tycho tested his table by measuring the refraction for Jupiter. He realized he had used an erroneous refraction table for Mars, and his results disappeared like a puff of smoke. Tycho's brilliantly conceived research programme had collapsed for reasons he could not fathom. Kepler, some years later, found the answer: the Solar System was at least three times larger than anyone had imagined.

In the intervening years, Tycho had moved to Prague, where he died, perhaps poisoned by the mercurial elixirs cooked up in his alchemical laboratory. The emperor procured the observational books for young Kepler, who found in them the goldmine of martian observations he needed to unlock the secrets of planetary astronomy. But this was no accident. They had been made with such marvellous accuracy for a purpose, by the man who defined what observing was all about. If we sing of Tycho, it is probably for the wrong reasons, but he was truly a pioneer who shaped the millennium.

Owen Gingerich is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

Cognitive ability and the light bulb

The road to the stars has also illuminated the human mind

Brian Aldiss

The arrival of the spaceship Conqueror into Arcopian space proves ironic. However, it provides us with an opportunity to look back on our distant predecessors and understand something of their combative and rickety societies.

Once the dead bodies had been cleared from the Conqueror, and preserved in our museums, mechs were dispatched to examine the ship as part of our phylogenetic record.

The ship was fitted with old-fashioned quantum computers. The Conqueror had left the old Solar System late in 2095. It carried 10,000 human embryos in cryogenic conditions, and several million embryos, similarly frozen, of terrestrial animal genera, together with numerous plant species. There were also 20 crew, supported by anti-tha-natonic drugs.

Technologists had designed the ship to accelerate to 12 per cent of the speed of light. By their computations, it was due to reach this system (where only two planets capable of supporting carbon-based life had been identified) in 196 years. The power source was a fusion engine.

In those rather primitive days, attention concentrated on the hardware. It was the bacteria on the Conqueror that brought about disaster, killing crew and embryos alike.

Subsequent advances in radiotelescope revealed no fewer than 15 planets orbiting the Arcopia main-sequence sun, of which five had environments suitable for life. In the Second Renaissance of the early twenty-second century, the spiritual order of God's Exiles perfected an ion drive and equipped another interstellar ship, Pilgrim. Pilgrim was launched from plutonian orbit in 2151. It carried with it the embryos of new species of animals, fruits and human beings. The entire journey was governed by quantors — God's Exiles did not inflict years of imprisonment on humans, as the Conqueror had done.

This journey took 138 years to complete. Thus the arrival was in 2289, two years before the Conqueror reached us, despite starting 56 years later.

In these improved drives we see symbols of the expansion of human consciousness. Everything is subject to change, and living things to evolutionary change, marking their passage through time. Study of the evolution of human consciousness was scarcely recog-

nized as a discipline until interstellar flight proved to speed conceptual processes. The necessity for understanding and dealing with totally new environments was responsible for this rapid acceleration in human mentation. A similar acceleration is recorded some 40,200 years ago in Europe, when fresh environments brought about a great expansion in the metaphors of art and sculpture — all of which represent an upward surge in cognitive ability.

To produce art or science is to experience a coming together of previously somewhat isolated faculties, which combine to make a greater whole. Another well-known example of such a quantal happening is the First Renaissance, a time of great advances in art, science, warfare and political management.

In describing these advances, twenty-second-century philosopher Almond Kunzel famously deployed an analogy between human consciousness and an antique, incandescent light bulb. Early consciousness could be likened to a 40-watt bulb — sufficient, dimly, to illumine a room, but insufficient to study details by. The First Renaissance marks a shift in brightness to 60 watts. Much more can be discerned, although illumination is not cast very far.

With the twentieth century, often referred to as the Savage Century, owing to its horrifying record of war, threats of war and genocide, the bulb brightens to 100 watts. Despite the savagery, humanity is for the first time developing a form of remote awareness (remware, as we know it)

to aid its exploration of all environments.

Those environments included, of course, the Solar System to which our predecessors were then confined, and also the human brain. The brain was almost completely mapped by the end of the Savage Century. With the ability to genetically engineer brain function, many irregularities — caused by the piecemeal nature of this organ as naturally evolved — were eradicated. Clearer thinking emerged. Humanity outgrew war.

We have now reached the stage of, in Kunzel's terms, the 1,000-watt brain. Our offspring are born with an understanding of fractals.

This great expansion of cognitive ability led to the new perception of the Universe as a series of contiguities, and to the construction, in the year 2162, of the photon drive. The fleet of ships launched in 2200 arrived here in the planetary system of Arcopia the following year.

Our culture was thus firmly established when the old ships of 2095 and 2151 arrived, fossils of a former time. They harbour, in orbits far from the planet on which humanity began, memories of bygone ages — long before there was even a light bulb to light our way. Records on these gallant old hulks demonstrate how, sadly, the human world once contained less order, less joy and less fulfilment than now.

Brian Aldiss is a novelist and critic. His latest book is a utopia, written in collaboration with Roger Penrose, entitled *White Mars, Or The Mind Set Free* (Little, Brown).



JACEY

A central control for cell growth

Alan J. Whitmarsh and Roger J. Davis

The identification of an unexpected function for the mitogen-activated protein kinase — the synthesis of DNA and RNA — opens up yet another route by which this multifaceted enzyme can influence cellular growth.

Cells respond to growth-promoting factors in two ways — by replicating their chromosomes and by increasing the expression of genes to form new proteins. Chromosome replication requires the production of new DNA, whereas the increase in gene expression requires synthesis of new ribonucleic acid (RNA). In both cases, however, increased production of nucleotides — the building blocks of RNA and DNA — is required.

On page 328 of this issue, Graves *et al.*¹ describe a hitherto unknown link between an enzyme that is critical for nucleotide synthesis and a growth-factor-induced signal-transduction pathway. Growth factors bind to specific receptors on the surface of the cell, triggering many signal-transduction pathways within that cell. One of these pathways leads to increased activity of an enzyme called the mitogen-activated protein (MAP) kinase, and activation of the extracellular-signal-regulated kinase (ERK) group of MAP kinases is necessary for cell growth². If ERK MAP kinase is activated constitutively — that is, switched on all the time — the result is unregulated growth of fibroblast cells^{3,4} and tumour formation in mice⁴. Conversely, thymocytes (immature T cells) isolated from mice that lack one member of the ERK family, ERK-1, grow more slowly in response to growth-promoting stimuli than do thymocytes isolated from normal mice⁵.

How does activation of ERK MAP kinase mediate increased cell growth? Graves *et al.*¹ show that, in response to growth signals, ERK MAP kinase controls the synthesis of nucleotides — the first step in the production of DNA and RNA. It does this by directly regulating the activity of an enzyme called carbamoyl phosphate synthetase II (CPS II), which catalyses the initial, rate-limiting step in the *de novo* synthesis of pyrimidine nucleotides⁶ (Fig. 1a). Both DNA and RNA are composed of pyrimidine (cytosine, thymine and uracil) and purine (adenine and guanine) nucleotides, which are distinguished depending on whether the common phosphate and sugar moieties are attached to purine- or pyrimidine-derived nitrogenous bases.

Graves *et al.* report that the activity of CPS II is increased after cells are exposed to the epidermal growth factor (which, as its name suggests, potently stimulates cell

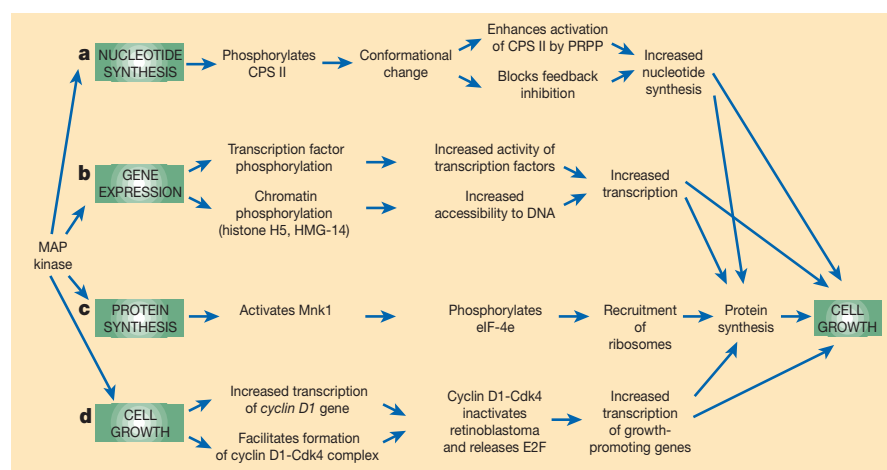


Figure 1 The mitogen-activated protein (MAP) kinase is central to cell growth. **a**, Graves *et al.*¹ show that MAP kinase increases the synthesis of pyrimidine nucleotides. Increased nucleotide synthesis is necessary for the new RNA and DNA synthesis that is required for cellular proliferation. **b**, MAP kinase also increases the activity of transcription factors to enhance gene expression, either by phosphorylating the transcription factors or by increasing the accessibility of the DNA. **c**, By phosphorylating the Mnk1 protein kinase, MAP kinase activates recruitment of protein-synthesis factors to messenger RNA. **d**, Activation of MAP kinase by growth-factor stimulation of cells leads to the increased expression of genes required for growth. MAP kinase increases cyclin D1 expression and induces the assembly of cyclin D1–Cdk4 complexes, which phosphorylate the tumour-suppressor protein retinoblastoma and cause the release of the E2F transcription factor. CPS II, carbamoyl phosphate synthetase II; PRPP, phosphoribosyl pyrophosphate; HMG, high-mobility group protein; eIF-4e, eukaryotic translation initiation factor-4E.

growth), or if the CPS II is incubated with ERK MAP kinase *in vitro*. Under these conditions, ERK MAP kinase catalyses the attachment of phosphate groups (a process known as phosphorylation) to CPS II at a particular amino acid (threonine 456), altering the conformation of CPS II. This in turn increases the activation of CPS II by the regulatory molecule phosphoribosyl pyrophosphate. The conformational change also blocks the feedback inhibition of CPS II activity by a product of pyrimidine-nucleotide synthesis, uridine triphosphate.

To show that ERK MAP kinase is directly responsible for these effects, Graves *et al.* inhibited the activation of ERK MAP kinase, or mutated the ERK MAP kinase phosphorylation site. In both cases, the increase in CPS II activity was blocked, suggesting that ERK MAP kinase regulates the activity of CPS II in a physiologically relevant way. Interestingly, the site at which ERK MAP kinase phosphorylates CPS II (threonine 456) is conserved in the highly related enzyme CPS I, which is involved in the synthesis of urea. This hints at yet another

function for ERK MAP kinase in a different biological process.

Another mechanism by which ERK MAP kinase mediates increased cell growth is the modification of transcription factors, which are involved in the transcription of DNA to form messenger RNA (mRNA). The ERK MAP kinase phosphorylates a group of transcription factors that turn on so-called rapid-response genes, the expression of which is required for growth⁷. The phosphorylated transcription factors then increase the rate at which these genes are transcribed, which in turn increases formation of the growth-inducing protein product (Fig. 1b).

As well as directly targeting the activity of transcription factors, ERK MAP kinase may also facilitate gene transcription by regulating the structure of chromatin. Chromatin is the packaged form of DNA found in chromosomes — the DNA is wrapped around proteins, including the histones and high-mobility group (HMG) proteins, meaning that transcription factors find it less accessible. So, for transcription to occur, the chromatin needs to be unravelled. Cells that lack

an enzyme called Rsk-2, which is activated by MAP kinase, cannot phosphorylate histone H3 (ref. 8) and so cannot unwrap the chromatin. Another MAP kinase-activated protein, MSK1, is also reported to phosphorylate histone H3 and HMG-14 (ref. 9). Although the exact function of these phosphorylation events is not known, they probably either directly improve the accessibility of DNA to transcription factors, or they allow the recruitment of chromatin-modifying enzymes.

The product of gene transcription, mRNA, is then translated into protein, and ERK MAP kinase is involved in this step too. Activation of ERK MAP kinase enhances translation by increasing the ability of a protein called the eukaryotic translation initiation factor-4E (eIF-4E) to recruit the protein-synthesizing ribosomes and other protein-synthesis-initiation factors to the mRNA. This effect may be mediated by the phosphorylation of eIF-4E by the MAP kinase-interacting kinase Mnk1, which is itself activated by ERK MAP kinase^{10,11} (Fig. 1c). So, activated ERK MAP kinase contributes to both increased gene expression and protein synthesis in cells stimulated with growth factor.

Finally, cell growth itself is influenced by ERK MAP kinase. Growth is tightly controlled by the activity of the cyclin-dependent protein kinase (CDK) family. The synthesis of a protein called cyclin D1, and its assembly with its partner Cdk4, is rate limiting for growth. Activation of the ERK MAP kinase signalling pathway increases transcription of the *cyclin D1* gene^{12,13}, and also facilitates the formation of an active cyclin D1–Cdk4 complex^{13,14} (Fig. 1d). This complex then phosphorylates and inactivates the growth-suppressing retinoblastoma protein. This inactivation, in turn, stops inhibition of a transcription factor called E2F, which can then increase the transcription of genes that promote growth and DNA replication.

It is clear from all of these studies that ERK MAP kinase has many functions in cell growth. But the study by Graves and colleagues¹ is the first evidence that ERK MAP kinase directly controls the synthesis of DNA and RNA. The constitutive activation of ERK MAP kinase seen in tumour cells may contribute to the formation of cancers by enhancing the activity of CPS II and increasing the supply of pyrimidine nucleotides. Indeed, CPS II activity is increased in many types of cancer cell¹⁵, probably giving them a selective advantage with respect to growth and DNA replication.

Further studies are required to establish whether ERK MAP kinase regulates additional steps in the synthesis of nucleic acids and to identify the members of the MAP kinase family that regulate these processes. Although the involvement of the ERK group

of growth-factor-activated MAP kinases has been established¹, the role of the p38 and the c-Jun N-terminal kinase (JNK) groups of stress-activated MAP kinases is not clear⁷. Finally, more work will be needed to examine the part played by MAP kinase in other processes that contribute to cellular proliferation.

Alan J. Whitmarsh and Roger J. Davis are at the Howard Hughes Medical Institute, Program in Molecular Medicine, Department of Biochemistry and Molecular Biology, University of Massachusetts Medical School, 373 Plantation Street, Worcester, Massachusetts 01605, USA.
e-mail: roger.davis@umassmed.edu

1. Graves, L. M. *et al.* *Nature* **403**, 328–332 (2000).
2. Lewis, T. S., Shapiro, P. S. & Ahn, N. G. *Adv. Cancer Res.* **74**, 49–139 (1998).
3. Cowley, S., Paterson, H., Kemp, P. & Marshall, C. J. *Cell* **77**, 841–852 (1994).
4. Mansour, S. J. *et al.* *Science* **265**, 966–970 (1994).
5. Pagès, G. *et al.* *Science* **286**, 1374–1377 (1999).
6. Jones, M. E. *Annu. Rev. Biochem.* **49**, 253–279 (1980).
7. Whitmarsh, A. J. & Davis, R. J. *J. Mol. Med.* **74**, 589–607 (1996).
8. Sassone-Corsi, P. *et al.* *Science* **285**, 886–891 (1999).
9. Thomson, S. *et al.* *EMBO J.* **17**, 4779–4793 (1999).
10. Waskiewicz, A. J. *et al.* *Mol. Cell. Biol.* **19**, 1871–1880 (1999).
11. Pyronnet, S. *et al.* *EMBO J.* **18**, 270–278 (1999).
12. Lavoie, J. N. *et al.* *J. Biol. Chem.* **271**, 20608–20616 (1996).
13. Ladha, M. H. *et al.* *Mol. Cell. Biol.* **18**, 6605–6625 (1998).
14. Cheng, M., Sexl, V., Sherr, C. J. & Roussel, M. F. *Proc. Natl Acad. Sci. USA* **95**, 1091–1095 (1998).
15. Aoki, T. & Weber, G. *Science* **212**, 463–465 (1981).

Quantum physics

Engineering decoherence

Wolfgang P. Schleich

The superposition principle lies at the very heart of quantum mechanics. As Paul Dirac stated in the first chapter of his classic textbook¹: “...any two or more states may be superposed to give a new state”. This principle sounds rather innocent but is the source of all the paradoxes of quantum mechanics, ranging from the Einstein–Podolsky–Rosen situation to Schrödinger’s cat. Superpositions can exist only in quantum systems that are free from external influences. But real quantum systems cannot stay isolated because a measurement requires interaction with an apparatus. If classical apparatus is used, which typically consists of a large number of atoms, then these atoms act as a reservoir. The interaction of the quantum system with the reservoir destroys the quantum-mechanical superposition. This decay from a superposition to a statistical mixture of states is called decoherence. On page 269 of this issue, Myatt *et al.*² from the National Institute of Standards and Technology in Boulder, Colorado, report a pioneering experiment that engineers decoherence.

Myatt *et al.* induce decoherence by coupling a single ion in a Paul trap to a reservoir that is under the experimenter’s control. Paul traps store individual charged particles such as ions and have been very successful in demonstrating quantum gates, the building blocks of a quantum computer. This is because laser cooling reduces the motional energy of the ion to the point where quantum effects become important. Myatt *et al.* engineer three different reservoirs, which allows them to confirm that the rate of decoherence scales exponentially with the size of the superposition, demonstrating why we never see macroscopic Schrödinger cats. This work is the culmination of several impressive experiments carried out

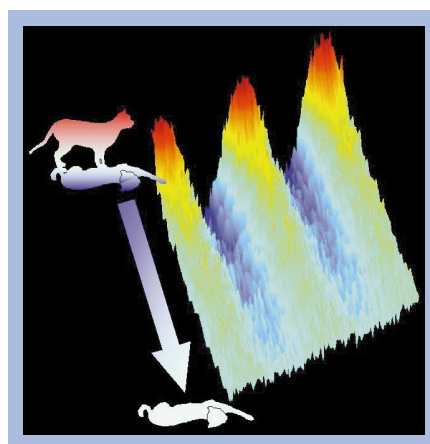


Figure 1 How the quantum world slides into the classical world. Myatt *et al.*² measure interference oscillations from a quantum superposition of a single atom being in two places at once. The oscillations gradually die out as the superposition gives way to the atom being in one definite place (either position). Alongside is Schrödinger’s cat, first in a superposition of states, then evolving to a definite state. Although the cat is ultimately dead in this case, alive would be equally as likely.

by Dave Wineland’s group at Boulder.

The ion in the Paul trap is an experimental realization of a quantum-mechanical harmonic oscillator. The most elementary model for a harmonic oscillator is a marble rolling under the influence of the gravitational field on a ruler bent into the shape of a parabola. In classical mechanics all energies are possible, but, as Max Planck showed in 1900, energy comes only in quantized units. Consequently, only discrete energies are allowed. In the language of the formal quantum mechanics developed by Werner Heisenberg and Erwin Schrödinger, the ion in the Paul trap can only have individual

energy eigenstates of motion. A few years ago the Boulder group prepared an ion in such a state³.

As well as energy eigenstates we can also have coherent states. To create them in the mechanical model we start from the ground state of the oscillator, that is, the marble resting in the minimum of the bent ruler. We suddenly displace the ruler and lower its centre, forcing the marble to move. In the case of an ion, such displacements are easy to obtain because an ion is sensitive to electric fields. By applying an electric field we can displace the ion from its ground state and create a coherent state. Again, the Boulder group created such states in earlier work³.

According to Dirac, the superposition of two coherent states of different displacements is allowed by quantum mechanics, but the superposition state has rather unusual properties: the ion is simultaneously at two distinguishable places. Schrödinger illustrated this paradox with his allegory of a cat that is simultaneously dead and alive. Indeed, such a state has been achieved experimentally using an ion in a trap⁴, or for an electromagnetic field in a cavity⁵.

The superposition state is not simply two individual peaks in phase space, but involves interference fringes, each being positive or negative contributions, halfway between the two peaks. (Phase space is a six-dimensional space involving three dimensions of position and three of momentum.) The spacing of the interference fringes depends on the separation of the two contributing peaks — the further the two peaks are separated, the closer the spacing. It is this interference in phase space⁶ that is not seen in classical physics: a classical state has two peaks but not the interference fringes.

When we let this quantum system interact with the outside world, the interference fringes disappear^{7,8} (Fig. 1); that is, the quantum system approaches the classical system by the process of decoherence. The speed at which the fringes disappear depends on the square of the separation of the contributing states. For the case of a cavity field, this decay towards the classical mixture has been observed by the group of Serge Haroche⁵. In Myatt and colleagues' experiment², three different environments have been used to study the decay of the interference fringes: an amplitude, a phase noise and a spontaneous emission reservoir. Whereas the first two examples are based on classical 'hot' reservoirs (noise at non-zero temperatures), the third is quantum noise at zero temperature.

For the amplitude reservoir the authors prepare a superposition of two coherent states and then apply a random electric field to the electrodes of the trap. This displaces the two coherent states; moreover, it also changes the phase of the fringes. For the phase reservoir they prepare the superposition of two energy eigenstates and then

apply a random but slow modulation of the steepness of the trap. This changes the phase of the superposition without changing the energy, leading to non-dissipative decoherence⁹. Both experiments confirm the exponential dependence of the decoherence on the square of the separation of the two states.

The Boulder group's third experiment addresses the question of engineering the reservoir to control decoherence. They follow a suggestion from Peter Zoller's group¹⁰ to use the internal structure of the ion. Two laser fields drive internal transitions in the ion, which, because of spontaneous emission, can decay from a given vibrational energy eigenstate to the ground state corresponding to zero temperature. In the absence of the laser field, the ion does not make the transition and is therefore not driven into this ground state. So, by choosing the appropriate laser fields, we can change the link with the reservoir and thereby engineer decoherence.

The work of Myatt *et al.*² has important consequences for the development of quantum computers. It is well known that

preventing decoherence is crucial to their successful operation. So engineering the reservoir, and therefore decoherence, may be one way to avoid complex error-correction schemes. Moreover, these experiments throw light on the border separating the classical from the quantum world. We have come a long way from Planck's discrete energy levels — now we can imagine experiments in which we could observe decoherence in real time. ■

Wolfgang P. Schleich is at the Abteilung für Quantenphysik, Universität Ulm, D-89069 Ulm, Germany.

e-mail: schleich@physik.uni-ulm.de

1. Dirac, P. A. M. *The Principles of Quantum Mechanics* (Clarendon, Oxford, 1958).
2. Myatt, C. J. *et al.* *Nature* **403**, 269–273 (2000).
3. Meekhof, D. M., Monroe, C., King, B. E., Itano, W. M. & Wineland, D. J. *Phys. Rev. Lett.* **76**, 1796–1799 (1996).
4. Monroe, C., Meekhof, D. M., King, B. E. & Wineland, D. J. *Science* **27**, 1131–1136 (1996).
5. Brune, M. *et al.* *Phys. Rev. Lett.* **77**, 4728–4731 (1996).
6. Schleich, W. P. & Wheeler, J. A. *Nature* **326**, 574–577 (1987).
7. Zurek, W. H. *Phys. Today* **44**, 36–44 (1991).
8. Giulini, D. *et al.* *Decoherence and the Appearance of a Classical World in Quantum Theory* (Springer, Heidelberg, 1996).
9. Bonifacio, R. *Nuovo Cimento* **114B**, 473–488 (1999).
10. Poyatos, J. F., Cirac, J. I. & Zoller, P. *Phys. Rev. Lett.* **77**, 4728–4731 (1996).

Neurobiology

Finding the lost target

Martin E. Schwab

All the sensations of our body — touch and heat, muscle tension, muscle ache and so on — reach the central nervous system (CNS) through the peripheral nerves, which enter the spinal cord as small, segmental nerve bundles, the spinal roots. When sensory nerve fibres are damaged in a peripheral nerve or a spinal root, they can regenerate over long distances. But when such regenerating fibres contact the spinal cord, they abruptly stop growing and seem unable to enter the CNS tissue. The study by Ramer *et al.*¹ on page 312 of this issue shows that specific neurotrophic factors (proteins that enhance the growth of nerve fibres and neuronal maturation during development) can influence the regenerating sensory fibres, allowing them to enter the spinal cord where they terminate and establish functional connections.

The supercomputer of our body, the CNS, requires a particular, specific environment: it is encapsulated in the bony structures of the skull and vertebral column, and it is shielded from the bloodstream by the blood–brain barrier. The motor output to the entire body leaves the CNS through the segmental ventral spinal roots, and the sensory input comes in through the corresponding dorsal roots. These delicate fibre bundles are prone to damage, often resulting

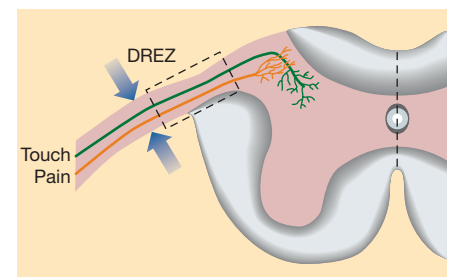


Figure 1 Cross-section of half the spinal cord with dorsal root and dorsal root entry zone (DREZ). Sensory primary afferents terminate in the DREZ and synapse in the dorsal horn in different areas corresponding to sensory modalities. Blue arrows indicate the damage to the dorsal root studied by Ramer *et al.*¹.

in massive pain. Strong tension on the roots — as could happen in, say, a motorcycle accident — can lead to the roots being torn from the spinal cord (root avulsion), leading to loss of sensation and paralysis.

Attempts to re-implant spinal roots surgically have met with only limited success. The main reason for this, in the case of sensory dorsal roots at least, seems to be the barrier function of the junction between the root and the CNS — the so-called dorsal root entry zone (DREZ; Fig. 1). The arrest of regenerating sensory fibres at the DREZ

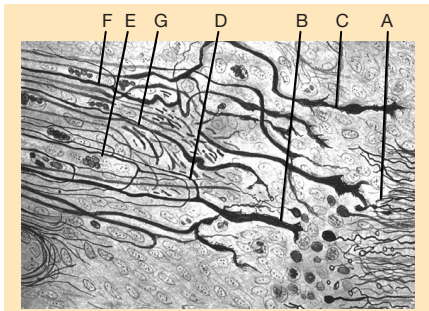


Figure 2 Picture from Ramón y Cajal (1928)². Interrupted dorsal root axons sprout (D–F) and form growth cones (B,C) but cannot enter the spinal cord (A). Ramer *et al.*¹ have now shown, however, that regenerating dorsal root fibres grow through the DREZ into the spinal cord under the influence of specific neurotrophic factors.

has been known for about a hundred years². Indeed, it has been studied as a model for the fact that nerve fibres can successfully regenerate in all parts of the peripheral nervous system, but not in the microenvironment of the spinal cord or brain (Fig. 2).

Several hypotheses have been proposed to explain this block to regeneration in the spinal cord and brain. These include the specific growth-promoting properties of peripheral-nerve glial cells (Schwann cells), as well as stop signals and growth-inhibitory cues associated with components of the CNS (in particular, the astrocytes of the DREZ and myelin-associated inhibitory molecules of the spinal white matter^{3,4}). In fact, by directly implanting dorsal roots into the grey matter (neuronal regions) of the dorsal horn of the spinal cord, or by bridging the DREZ, some growth of sensory fibres into the spinal cord has been seen⁵.

Ramer *et al.*¹ have now used a panel of neurotrophic factors to stimulate the regeneration of damaged sensory fibres. Surprisingly, these fibres overcome the inhibitory influences of the DREZ: they grow into the spinal cord and form branches in specific parts of the dorsal horn, their normal target area (see Fig. 1 on page 313).

Neurotrophic factors have many complex effects on responsive nerve cells, depending on the age of the neurons and the context in which the factors are presented. For example, developing dorsal root ganglion neurons with different functions (such as stretch receptors, touch-sensitive neurons, temperature sensors or pain sensors) respond to, and depend on, particular neurotrophic factors, especially those belonging to the neurotrophin family — nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3) and neurotrophin-4 (NT-4). Neurotrophic factors can also enhance the growth of fibres during development, in culture, and even within the CNS after damage to adult CNS

fibre tracts^{3,6,7}. In addition, they can reprogramme the response of neuritic growth cones to, say, repulsive or inhibitory factors⁸.

Ramer *et al.*¹ infused rat brain with NGF, BDNF, NT-3 or glial-cell-line-derived neurotrophic factor (GDNF), and saw that interrupted axons grew through the DREZ into the spinal cord and its dorsal horn. However, the different growth factors had unequal effects — whereas GDNF had a strong growth-enhancing effect for different types of sensory axons, NT-3 affected mainly the large-calibre mechanoreceptors, and small-calibre pain and temperature fibres responded mainly to NGF. This specificity is in line with the developmental effects of these factors and the distribution of the corresponding receptors.

The next question is whether, when damaged nerve fibres regrow in the adult nervous system, they can find their targets and establish functionally meaningful connections. Little information is available in the CNS, although some behavioural improvements have been observed in studies of experimentally induced regeneration after damage to the spinal cord. This suggests that the mechanisms for target recognition and stabilization of functionally meaningful connections should be present^{7,9,10}.

Ramer and colleagues searched for evoked electrical activity in postsynaptic cells of the dorsal horn of their neurotrophin-treated, dorsal-root-lesioned animals. They found clear evidence for functional re-innervation of these spinal neurons by the regenerating fibres — the modality- and fibre-type-specific innervation profile of the dorsal horn was restored. The authors then tested functional reconnection by two behavioural reflexes: the protective (withdrawal) responses to pressure and to water at 49 °C applied to the forepaw. Damage to the dorsal root shifted the withdrawal responses greatly in both tests, indicating that the rats could not register the disturbing or painful stimuli. But recovery of both responses was seen within 2–3 weeks in the GDNF- and NGF-treated animals. Both of these neurotrophic factors influence pain- and temperature-sensitive neurons. So these results show that regenerating axons have contacted post-synaptic target neurons in the spinal cord in a functionally correct manner.

The factors and mechanisms that prevent or enhance the regeneration of damaged adult peripheral and central axons are now rapidly being unravelled. The demonstration that regrowing fibres can recognize target areas and form functionally meaningful connections (meaningful on the level of behaviour and the organism) is an encouraging step. Target-recognition signals may, in fact, be expressed throughout life in circuits that retain some structural plasticity. Strengthening of functionally meaningful synaptic connections (and retraction of



100 YEARS AGO

Is New Zealand a Zoological Region?

Will you allow me to make one remark on the letter of Mr. H. Farquhar, advocating an affirmative answer to the above question. It is this: Throughout the whole argument there is an assumption which vitiates it, namely, that the amount of resemblance of the New Zealand fauna to that of Australia is what alone determines its resemblance to that of the Australian Region. Apparently, Mr. Farquhar does not believe that New Caledonia and the New Hebrides belong to the Australian Region, otherwise he would not adduce the fact of the land-shells of New Zealand being related to those of the above-named islands as an argument in his favour; and if these are omitted, then must New Guinea be also omitted. And if Australia by itself is to become a "Zoological Region," New Guinea and its surrounding islands must be also a "Region," the Central Pacific Islands another, and the Sandwich Islands yet another! This indicates the difficulties that arise if the Australian Region, as originally defined by Dr. Sclater and myself — and which I still hold to be far more natural than any subdivision can make it — be rejected.

Alfred R. Wallace.

From *Nature* 18 January 1900.

50 YEARS AGO

A new insect-proof packaging material which is cheap and very easy to handle has been developed by the Pest Infestation Laboratory of the Department of Scientific and Industrial Research. There is considerable wastage in packaged foods, resulting from the penetration of insects from outside. It was found that a layer of sand-paper formed an effective barrier, but did not prove practicable in use. Paper and corrugated cardboard impregnated with D.D.T. were unsuccessful because the insects bored straight through and so did not remain long enough in contact with the insecticide to pick up a lethal dose. The final solution was to use cellulose wadding, several layers thick, impregnated with D.D.T. When faced with this material the insect, after getting through the first layer, wanders in all directions and so takes up enough D.D.T. to be killed. Not one insect has ever got through this material in the course of numerous laboratory tests.

From *Nature* 21 January 1950.

non-functional ones), based on mechanisms similar to those that operate during development, may be another important mechanism. In the specific case of spinal root avulsion, Ramer and colleagues' results should stimulate both preclinical and clinical research. The goal here would be to re-implant and re-connect disrupted spinal roots, making paralysed and numb limbs functional again. ■

Martin E. Schwab is in the Department of Neuromorphology, Brain Research Institute, University of Zürich, and the Swiss Federal Institute of Technology Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland.

e-mail: schwab@hifo.unizh.ch

1. Ramer, M. S., Priestley, J. V. & MacMacon, S. B. *Nature* **403**, 312–316 (2000).
2. Ramón y Cajal, S. *Degeneration and Regeneration in the Nervous System* (Hafner, New York, 1928).
3. Schwab, M. E. & Bartholdi, D. *Physiol. Rev.* **76**, 319–370 (1996).
4. Fitch, M. T. & Silver, J. *Cell Tissue Res.* **290**, 379–384 (1997).
5. Siegal, J. D., Klot, M., Smith, G. M. & Silver, J. *Exp. Neurol.* **109**, 90–97 (1990).
6. Schnell, L., Schneider, R., Kolbeck, R., Barde, Y.-A. & Schwab, M. E. *Nature* **367**, 170–173 (1994).
7. Grill, R., Murai, K., Blesch, A. & Gage, F. H. *J. Neurosci.* **17**, 5560–5572 (1997).
8. Cai, D., Chen, Y., De Bellards, M. E., Tang, S. & Filbin, M. T. *Neuron* **22**, 89–101 (1998).
9. Bregman, B. S. *et al. Nature* **378**, 498–501 (1995).
10. Li, Y., Field, P. M. & Raisman, G. *Science* **277**, 2000–2002 (1997).

Atmospheric chemistry

Better budgets for methyl halides?

James H. Butler

Methyl bromide (CH_3Br) and methyl chloride (CH_3Cl) both originate in large part from natural sources. These gases can reach the stratosphere where their halogen atoms, released through photolysis, catalytically destroy ozone. Three papers in this issue^{1–3} (pages 292–301) may help to close large gaps in the calculated budgets of methyl halides in the atmosphere. They also open up possibilities for improving our understanding of the fluxes of other naturally produced, halogenated gases.

Over the past two decades, attention in atmospheric chemistry has centred on halocarbons that result from human activities — the chlorofluorocarbons (CFCs), halons (CBrF_3 , CBr_2ClF) and chlorinated solvents (CH_2Cl_2 , CCl_4 , CH_2Cl_2 , CHCl_3), and their replacements, the hydrochlorofluorocar-

bons (HCFCs) (Fig. 1). Halogen atoms from these compounds also attack stratospheric ozone but the gases are entirely anthropogenic in origin. Only in the past five or six years have the methyl halides, which have both natural and human sources, received similar attention.

Methyl bromide is the single largest carrier of bromine to the stratosphere, where, on a per-atom basis, it is 50–60 times more effective than chlorine in depleting ozone⁴. Although natural sources dominate the methyl bromide budget, there is a sizeable anthropogenic flux to the atmosphere through its use as a fumigant. Its industrial production is due to be phased out, largely because of its high ozone-depletion potential. Methyl chloride, on the other hand, is the most abundant chlorine-containing compound in the atmosphere, contributing over 15% to the total tropospheric burden of organic chlorine. Its sources are believed to be largely natural and there is some evidence that it was present at over 90% of today's levels during the early twentieth century⁵. Both methyl halides have lifetimes of around a year, making them much shorter-lived than the CFCs, solvents and halons currently banned by international agreement. Nevertheless, their large fluxes into the atmosphere mean that they reach the stratosphere, where they become involved in ozone depletion.

The issues facing scientists studying these two compounds are largely to do with unbalanced atmospheric budgets. In the *Scientific Assessment of Ozone Depletion: 1998* (ref. 4), sources for only 60% of the sinks for methyl bromide could be accounted for. The estimates for methyl chloride are more uncertain, but again only about half to two-thirds of the sinks can be balanced with sources.

From previous work^{6–8}, it seems that the missing sources of the two gases are not

oceanic. For the most part, the ocean is undersaturated in methyl bromide and is insufficiently supersaturated in methyl chloride to explain more than a small percentage of its total flux to the atmosphere. This has led to several studies to determine whether methyl halides are released in significant amounts elsewhere. It appears that they are. Two of the new papers, those by Yokouchi *et al.*¹ and Rhew *et al.*², identify high emissions of methyl halides from terrestrial-coastal ecosystems. The third, by Keppler *et al.*³, introduces evidence for an abiotic mechanism of organic halide production in organic-rich soils.

The work of Yokouchi *et al.*¹ was observational in nature. It involved monitoring the atmosphere at fixed sites in the western Pacific Ocean, sampling on cruises running from Antarctica to Alaska, and making spot checks of coastal environments. Their results show higher concentrations and higher variability in atmospheric methyl chloride in coastal areas. They also found that levels of methyl chloride were correlated with levels of α -pinene, a compound that

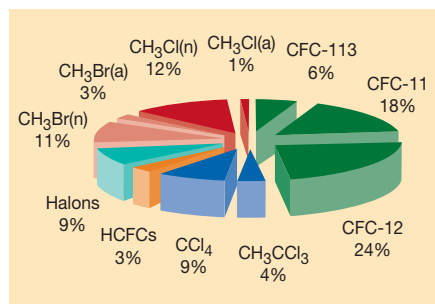


Figure 1 Equivalent chlorine in ozone-depleting halogenated gases. Data are for 1999 concentrations in the troposphere¹². (n) and (a) denote estimates of natural and anthropogenic contributions, respectively. Halogens in the two methyl halides make up about a quarter of the equivalent chlorine from persistent organic compounds in the atmosphere. (Equivalent chlorine is the total number of chlorine atoms plus a weighting factor (~50) multiplied by the total number of bromine atoms contributed by these compounds.)

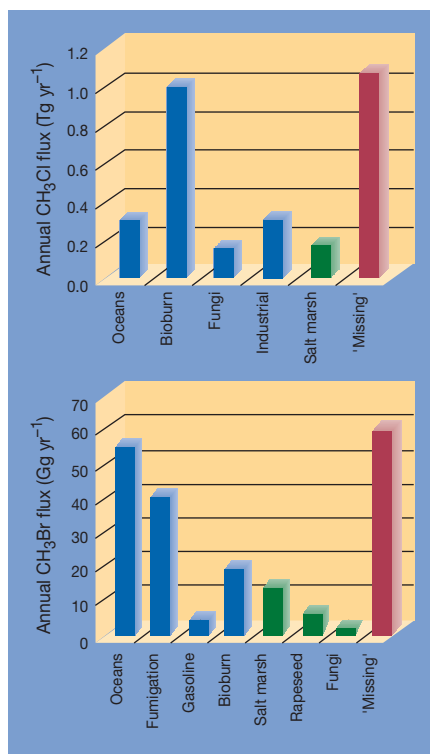


Figure 2 Sources of atmospheric methyl chloride and methyl bromide. Sources identified since the *Scientific Assessment on Ozone Depletion: 1998* (ref. 4) are shown in green. Budget deficits, calculated as the sum of all identified sinks minus the sum of all identified sources, are shown on the right. Salt marsh estimates are from Rhew *et al.*². Estimates of fluxes from rape seed and fungi are from refs 10 and 11. (Although the oceanic source of CH₃Br is large, its net flux is actually negative because the oceanic sink, which is included in the sink calculations, is even larger^{6,7}.)

derives specifically from terrestrial plants and has an atmospheric lifetime of only hours. The implication is that the source of the methyl chloride was both recent and terrestrial. From their cruise data, mainly taken in the western Pacific, Yokouchi *et al.* also report that concentrations of methyl chloride are increased in the tropics relative to higher latitudes. This, they claim, is consistent with their view that the gas has a strong, terrestrial-coastal source. It is also in line with other data suggesting⁹ that methyl chloride production is especially high in the tropics.

Rhew *et al.*² also observed strong methyl halide fluxes from a coastal environment. In a field study of two coastal marshes in southern California, they recorded increasing concentrations of methyl bromide and methyl chloride in experimental chambers. They found that the fluxes of the two gases varied with inorganic halide availability, species composition, time of day, and season. Their maximum observed fluxes were higher than those from freshwater wetlands by factors of 500–750. They also report a strong correlation between emissions of methyl bromide and those of methyl chloride, and suggest that the gases emanate from similar processes (although methyl bromide production may be favoured). Of particular interest is the finding that salt marshes, which make up less than 0.1% of the global surface area, could be producing 10% of the global emissions of the two gases.

Both Yokouchi *et al.* and Rhew *et al.* invoke biologically mediated reactions to produce the methyl halides. But that may not be the whole story, as the paper by Keppler *et al.*³ shows. These authors describe results from a series of controlled laboratory experiments, and offer an entirely new suggestion for the production of methyl halides and other organic halogens in the environment — that of an abiotic sequence of oxidation–reduction reactions in an organic medium. The mechanism involves the oxidation of organic matter at the expense of a cation, notably Fe^{3+} , a halide ion being methylated in the process.

These reactions could be ubiquitous, but they are heavily dependent on the availability of the right source materials. Keppler *et al.* stop short of providing global estimates from their results, but it is clear that such a mechanism could generate large amounts of methyl bromide, methyl chloride and other low-molecular-weight organohalogenes. It remains to be seen whether halogenated gases produced in this way are distributed widely enough to have a global impact, or even whether they reach the atmosphere before being degraded.

Together with other findings on crops and fungi^{10,11}, the three new papers^{1–3} provide further evidence that terrestrial systems,

especially coastal areas where halide ions are in ample supply, can make significant contributions to the global atmospheric burden of halogenated organic gases (Fig. 2). More work needs to be done — indeed some is under way — to understand the distribution and magnitude of these fluxes, and the biochemical and inorganic mechanisms that could be driving them. With such knowledge, we can perhaps begin to predict what shifts in the balance of methyl halide production and degradation might occur in the face of global change.

James H. Butler is at the Climate Monitoring and Diagnostics Laboratory of the National Oceanic and Atmospheric Administration, 325 Broadway, Boulder, Colorado 80303, USA.

Evolutionary biology

Protamine wars

Andrew G. Clark and Alberto Civetta

Proteins that evolve with unusual speed pique our interest because the mere rapidity of their change may be a clue to their function and adaptive significance. Proteins involved in sexual reproduction are particularly exciting because of their potential function in determining the success of sperm and eggs. Indeed, such 'sex-related' proteins do, generally, show accelerated evolution^{1,2}. A report by Wyckoff *et al.*³ on page 304 of this issue now highlights the accelerated evolution of the genes encoding a group of proteins called the protamines, which are associated with male reproduction. The authors suggest that this rapid evolution is driven by sexual selection, and the story involves some interesting twists.

The rate of change of a protein's molecular sequence may accelerate if the protein faces relaxed functional constraints or if natural selection drives the changes. In both cases, the result is that the rate of amino-acid substitution increases relative to that seen in a conservatively maintained protein. If all single-base changes in DNA were equally likely, about three-quarters of them would result in an amino-acid change in the encoded protein. But in most genes, a much lower fraction of changes result in amino-acid substitutions, and an excess rate of amino-acid change is generally ascribed to natural

e-mail: jbutler@cmdl.noaa.gov

1. Yokouchi, Y. *et al.* *Nature* **403**, 295–298 (2000).
2. Rhew, R. C., Miller, B. R. & Weiss, R. F. *Nature* **403**, 292–295 (2000).
3. Keppler, F., Eiden, R., Niedan, V., Pracht, J. & Schöler, H. F. *Nature* **403**, 298–301 (2000).
4. Kurylo, M. J. *et al.* in *Scientific Assessment of Ozone Depletion: 1998* (ed. Ennis, C. A.) (World Met. Org., Geneva, 1999).
5. Butler, J. H. *et al.* *Nature* **399**, 749–755 (1999).
6. Lobert, J. M. *et al.* *Science* **267**, 1002–1005 (1995).
7. Lobert, J. M., Yvon-Lewis, S. A., Butler, J. H., Montzka, S. A. & Myers, R. C. *Geophys. Res. Lett.* **24**, 171–172 (1997).
8. Moore, R. M., Groszko, W. & Niven, S. J. *J. Geophys. Res.* **101**, 28529–28538 (1996).
9. Khalil, M. A. K. & Rasmussen, R. A. *Atmos. Environ.* **33**, 1305–1321 (1999).
10. Gan, J., Yates, S. R., Ohr, H. D. & Sims, J. J. *Geophys. Res. Lett.* **25**, 3595–3598 (1998).
11. Lee-Taylor, J. M. & Holland, E. A. *J. Geophys. Res.* (in the press).
12. Montzka, S. A. *et al.* *Nature* **398**, 690–694 (1999).

selection^{4–6}. The high rate of amino-acid substitution in protamines⁷ led to the conclusion that they deviate from a strictly neutral model, and that some form of natural selection must be driving the change^{8,9}. To assess the importance of this finding, we need to consider the function of protamines.

During sperm maturation in many animals, protamines replace histones as the protein that maintains DNA in a compacted form in chromosomes. Like histones, protamines can associate strongly with DNA because they are highly basic, having arginine residues at about 50% of their amino-acid positions. In mammals, protamines also have between six and nine conserved cysteine residues, which form disulphide bridges within and between protamine molecules and maintain a tightly compacted structure¹⁰.

Alignment of protamines from different primates (Fig. 1) reveals that, although the proteins have uniformly high levels of arginine, the positions of the arginine residues vary widely. In fact, in both protamines P1 and P2, arginine is present at over 60% of the positions where there are sequence differences between various species of primate. So could the patterns of excess amino-acid change, which Wyckoff *et al.*³ argue are caused by sexual selection, instead be caused

```
MARYCCRSQSRSRYRQRQSRRRRRRRSCQTRRRAMCCRPAYRPRCRRH Human
.....C.....K.Q.....RRS.M.R... Chimpanzee
.....C.....T.....Q.....RRN.L.R.K. Gorilla
.....Q.CC.R...CH...C.....RR...L... Orangutan
      * * * * *
```

Figure 1 Alignment of the protamine P1 (PRM1) amino-acid sequences from human, chimpanzee, gorilla and orangutan. There is an unusually high incidence of changes at sites (asterisks) that encode the highly basic amino-acid arginine (R). High arginine content is necessary for DNA-binding ability, and the protein seems to tolerate considerable variation in the arrangement of its arginine residues.

by natural selection, which maintains a high overall arginine content but tolerates the gain and loss of arginine residues in different positions?

Perhaps, but which of these competing theories do the data best support? A key part of Wyckoff and colleagues' argument³ is the supposition that altered protamines affect sperm morphology. Protamines are necessary for male fertility — infertile men often have abnormal ratios of protamines P1 and P2, and men with no protamine P2 are sterile¹¹. But these are examples of quantitative differences in the levels of protamine expression, rather than the qualitative changes in protamine sequences.

There is too little variation among human protamine genes to see any difference in fertility associated with an altered protamine sequence. But we can turn the question around and ask whether there are any unusual features in the protamines of men whose sperm have abnormal morphology. Only some cases of round-headed sperm syndrome are associated with abnormal protamine levels¹², suggesting at best a weak causal link. More generally, the degree of condensation of chromosomes in sperm does not seem to be associated with rates of fertilization, implantation or pregnancy¹³.

Perhaps the most definitive way to test the function of a gene is to insert the gene *de novo* into an organism and look for effects. When Stewart *et al.*¹⁴ expressed the human protamine gene cluster in mice they found normal testis-specific expression, sperm morphology and fertility. Likewise, when an avian protamine called galline was expressed in the mouse¹⁵ there was abnormal chromatin condensation, yet many of the resulting spermatozoa were functionally normal and most of the mice had normal fertility. These transgenic experiments are not definitive proof of functional equivalence, but they also do not support the idea that the rapid sequence divergence is caused by selection for functional divergence. Instead, they indicate that the main feature that is important to the function of protamines is their ability to bind DNA and compact chromatin.

If the only attribute driving the evolution of protamines was their ability to bind DNA, would we expect to see the observed significant excess of amino-acid changes and rapid evolution^{3,9}? We have done a simulation of a protein sequence with natural selection, maintaining an optimum arginine content of 50% but ignoring the positions of those arginines, and find excess amino-acid substitution. So, although the statistics reported by Wyckoff *et al.* clearly show that the changes in protamines are not simply random, they do not help to resolve the two contrasting explanations.

It seems that the only way out of this impasse is to make reciprocal transgenic animals and test the function of the resulting

sperm. Ideally these experiments would be done by gene replacement, and careful controls would be needed to assure that appropriate expression levels were achieved. In any case, rigorous analysis of the cause of evolutionary changes in human genes may often run into a brick wall because we cannot do the definitive experiment. Barring this, analytical approaches like that of Wyckoff *et al.*³ may provide key insights to the evolutionary forces at play on human genes.

Andrew G. Clark and Alberto Civetta are at the Institute of Molecular Evolutionary Genetics, Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

e-mail: c92@psu.edu

1. Civetta, A. & Singh, R. S. *Mol. Biol. Evol.* **15**, 901–909 (1998).
2. Civetta, A. & Singh, R. S. *Genome* **42**, 1033–1041 (1999).
3. Wyckoff, G. J., Wang, W. & Wu, C.-I. *Nature* **403**, 304–309 (2000).
4. Hughes, A. L. & Nei, M. *Nature* **335**, 167–170 (1988).
5. Endo, T. *et al. Mol. Biol. Evol.* **13**, 685–690 (1996).
6. Schmid, K. J. *et al. Genetics* **153**, 1717–1729 (1999).
7. Queralto, R. *et al. J. Mol. Evol.* **40**, 601–607 (1995).
8. Rooney, A. P. & Zhang, J. *Mol. Biol. Evol.* **16**, 706–710 (1999).
9. Rooney, A. P. *et al. Mol. Biol. Evol.* **17**, 278–283 (2000).
10. Retief, J. D. *et al. Mol. Biol. Evol.* **12**, 708–712 (1995).
11. de Yebra, L. *et al. J. Biol. Chem.* **268**, 10553–10557 (1993).
12. Carrell, D. T. *et al. Fertil. Steril.* **71**, 511–516 (1999).
13. Hammadeh, M. E. *et al. Hum. Reprod.* **14**, 363–367 (1999).
14. Stewart, K. S. *et al. Fertil. Steril.* **71**, 739–745 (1999).
15. Maleszewski, M. *et al. Biol. Reprod.* **58**, 8–14 (1998).

erratum

Figure 1 of Victoria Lundblad's News and Views article "Telomeres: A tale of ends" (*Nature* **403**, 149; 2000) was printed incorrectly in some editions. The correct version is shown below.

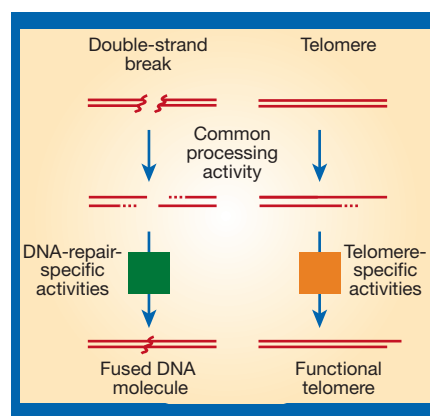


Figure 1 Proposed set of common events at DNA double-strand breaks and telomeres. Evidence is growing for the idea that these two pathways are partially linked. Ahmed and Hodgkin¹ show that the *mortal germline-2* (*mrt-2*) gene, which is required for telomere replication, is also involved in monitoring DNA for double-strand breaks. The idea is that processing to resect one strand of a telomere or double-strand break could be resolved by activities that are specific for telomere maintenance (such as telomerase) or DNA repair (such as a DNA ligase).

1. Ahmed, S. & Hodgkin, J. *Nature* **403**, 159–164 (2000).

Daedalus

States of non-mind

Why are we conscious? Evolution has been blamed. For some unexplained reason, conscious creatures were fitter than ones which behaved the same way but without consciousness; and Darwin did the rest.

If so, says Daedalus, then consciousness must be represented somewhere on the genome. Creatures which are conscious should have a set of genes absent in those which are not conscious. And consciousness is remarkably unitary. It can be abolished by many simple molecules (anaesthetics), leaving other bodily systems untouched. So it is probably coded for by one, or just a few, genes. How to identify them?

Daedalus recalls the strange syndrome of alcoholic 'palimpsest'. The alcoholic has absolutely no memory of some past episode, even though at the time he did not appear drunk. Actors have given brilliant performances, surgeons have conducted delicate operations that they would never later recall, in alcoholic palimpsest.

Daedalus reckons that palimpsest is not a failure of memory, but of consciousness. Alcohol, itself a considerable anaesthetic, erases the consciousness of its victim. It leaves him as a perfectly functional human robot, seemingly normal and responsive, but in fact with no internal awareness.

So DREADCO biochemists are conducting delicate tests on advanced alcoholics, looking for hormones and neurotransmitters whose concentration correlates with the state of palimpsest. They hope to work out where these compounds come from, on what brain sites they bind or inhibit binding, and what proteins underlie their synthesis. Standard genetic detective work should then reveal the crucial genes of consciousness.

A major biological conundrum will thus be resolved. By seeking these genes in other creatures, the DREADCO team will discover which of them is conscious. Caring experimental biologists will rush to learn the results. Are experiments on mice, or frogs, or fish, or beetles, cruel? Or are these creatures merely unconscious chemical mechanisms? Genetics will give the answer. Daedalus even hopes to grant awareness to lower species by implanting the crucial genes in them, or deprive higher ones of it by deleting them. But how to tell if the operation has worked? Daedalus will induce alcoholism in the test creatures, and look for episodes of palimpsest. **David Jones**

The Further Inventions of Daedalus (Oxford University Press), 148 past Daedalus columns expanded and illustrated, is now on sale. Special *Nature* offer: m.curtis@nature.com

Obituary

Robert A. Swanson (1947–99)

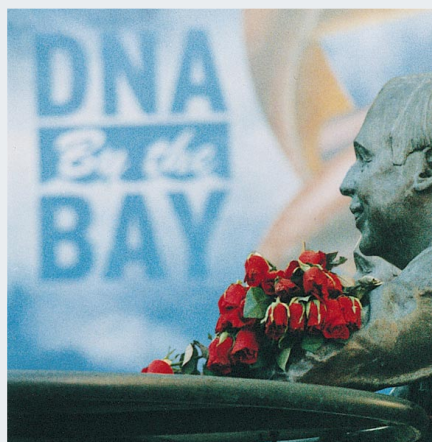
On 6 December, the founder of Genentech, Robert Swanson, succumbed to brain cancer. He was only 52, and with his death the world has lost one of its visionaries — a man who is widely considered to be the father of the biotechnology industry, and who a year ago was included as one of the few living people on a list of the millennium's most influential individuals.

Swanson grew up in Florida and was educated at the Massachusetts Institute of Technology, obtaining degrees in chemistry and business in 1970. He started his professional career as a venture capitalist with Citicorp in New York, then in 1973 moved west to San Francisco and began a lifelong fascination with molecular biology. He was in the right place at the right time, as the pioneering gene-cloning experiments of Herbert Boyer, Stanley Cohen and colleagues were being carried out only minutes from his office.

In 1976, he arranged a fateful meeting with Boyer — a meeting now commemorated by a life-sized statue of the two men in the Genentech research complex (Swanson's is shown in the picture here). After several hours and a few beers, they agreed to team up to create Genentech, the world's first biotechnology company. The upshot was an industry that is now worth more than \$100 billion and is responsible for around 80 approved drugs that benefit millions of people.

Although Boyer laid the scientific foundation for the company and set its early research direction, he remained at the University of California, San Francisco (UCSF). So it was the 28-year-old Swanson who ran Genentech. Initially, this meant raising enough money to fund proof-of-principle projects at UCSF and the City of Hope Medical Center. In 1977 these groups demonstrated that bacteria could be coaxed into making the human protein somatostatin. Swanson then set his sights on human insulin, a previously unavailable pharmaceutical, and rented space in a warehouse to build Genentech's labs.

This transition, from a 'paper' corporation that supported research in academic labs to a real company, was a critical juncture for Genentech. Swanson now had to convince young postdocs to forgo the standard academic career path and to join his quest to create medicines through the power of a new technology. He succeeded because he was able to generate his own conviction and excitement in others, as we know only too well — we



Venture capitalist and the Pied Piper of biotechnology

were two of his earliest recruits. Remember that at the time genetic technology had limited funding, and even leading molecular biologists thought it was neither commercially nor intellectually viable.

Swanson was a businessman, but he never gave up hope in the science or the scientists during the early days. It was Swanson who was smart enough to hire the best people and give them an environment that allowed them to make a difference. Swanson also reminded us why we were doing it — for the patients — and he challenged employees with the ultimate personal motivation, "Would you put this drug into your own child's body?"

And it was Swanson who made all that hard work a lot of fun. He led the celebration as each milestone passed, including the early successes with drugs such as insulin, growth hormone, interferon and factor VIII to name a few. In 1987, when the US Food and Drug Administration approved Activase (tissue plasminogen activator) for the treatment of heart attacks, the festivities included, with permission from nearby San Francisco International Airport, a huge firework display that caused a temporary disruption of air traffic.

Swanson and Boyer started a company, but they also nurtured a unique culture at Genentech — one that departed radically from the standard in the pharmaceutical industry. First, there was very little hierarchy, a structure that encouraged employees to take risks and that paid huge dividends in scientific thinking and productivity. Second, there was the need to

attract the best postdocs and scientists from academia.

Twenty years ago it was considered heresy for top academic scientists to join industry. Swanson and Boyer attacked the reluctance at its source, which was the inability of pharmaceutical companies to see the connection between basic research and ultimate commercial profitability. Boyer had made it clear that, to recruit top scientific talent, Genentech would have to allow them to publish their discoveries, and promptly. This policy gave Genentech a recruiting advantage and forced many other companies to establish a similar approach. Part of Swanson's legacy is the number of seminal

publications in biology produced by scientists employed by industry.

Genentech was again an innovator in 1980 when it became the first biotech company to sell stock to the public. This was quite an education, for many of us had no idea what it meant to 'go public'. Employees were told that the stock might be floated at \$20 per share. The deal was finally priced at \$35 per share and the first shares traded at \$88. Swanson and Genentech had created business history long before the word 'Internet' had become common currency. Furthermore, we molecular biologists now understood that it was possible not only to create life-saving drugs, but also to be well compensated for our work. We felt we had the best jobs in the world.

Swanson was chief executive officer of Genentech until 1990, then served as chairman until his retirement in 1996. That year he joined the board of Tularik, a private biotech company specializing in gene regulation, as chairman. He also returned to his roots by starting his own venture-capital firm, and continued to enjoy helping young scientists and businessmen turn their ideas into start-up companies. However, Bob's family remained his top priority. He is survived by his wife, Judy, and daughters Katie, 16, and Erica, 11.

Bob Swanson was dedicated to generating drugs that would save lives. In following that aim, he also made people believe, like him, that the seemingly impossible was indeed possible. It is that spirit which will live on.

David V. Goeddel and Arthur D. Levinson

David V. Goeddel is at Tularik, Inc., 2 Corporate Drive, South San Francisco, California 94080, USA. e-mail: dvg@tularik.com

Arthur D. Levinson is at Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA. e-mail: alevinson@gene.com

Ultrasonic hearing in nocturnal butterflies

Hedyliids have ultrasound-sensitive ears on their wings to help them avoid bats.

Ultrasonic hearing is common in moths, which rely on it for defence and communication^{1,2}, but it has never been demonstrated in butterflies. Here we describe a new type of ultrasound-sensitive ear that we have discovered in an unusual group of nocturnally active, neotropical butterflies, the Hedyloidea. Hedyliids have ears on their wings and respond to ultrasound by making flight manoeuvres to avoid bats. On the basis of phylogenetic and comparative anatomical evidence, we propose that hearing in Lepidoptera and day flight in butterflies both result from an intense selection pressure imposed by echolocating bats more than 50 million years ago.

Diurnality has traditionally been used to distinguish butterflies from moths, and underlies a major difference in the sensory worlds of these two groups. Butterflies are predominantly diurnal and rely on a well-developed visual system for communication and predator detection³. Although it has been proposed that some butterflies have low-frequency hearing^{4,5}, ultrasonic hearing is unknown for this taxon. In contrast, most moths have tympanal ears that function primarily to detect echolocation calls used by insectivorous bats to locate and track their prey¹.

The Hedyloidea, which have previously been identified as moths, are now believed to be the closest relatives, and possibly the 'living ancestors', of modern-day butterflies (Papilionoidea, Hesperoidea)^{6,7}. The forewing modifications of Hedyloidea have been suggested to function in hearing or scent production⁶, but until now they had not been studied in living individuals. Because hedyliids fly at night and are therefore exposed to bats, we predicted that they should have ultrasound-sensitive ears and exhibit bat-avoidance behaviour.

Our anatomical investigations of *Macrosoma heliconiaria*, collected live on Barro Colorado Island, Panama, have revealed a tympanal ear⁴ of intriguing design. The eardrum (tympanic membrane) is nestled within a tympanic cavity at the narrow end of a canal that resembles the pinna of a rabbit ear (Fig. 1a,b). The thin (1 μ m) membrane stretches over a large air-filled chamber and has three vibration-sensitive chordotonal sensory organs attached to separate regions of its inner surface, similar to the frequency-discriminating ears of acridid grasshoppers⁸. The ears move with the wing and so may be efficient at localizing sounds during flight. Comparative anatomical studies indicate that ears are

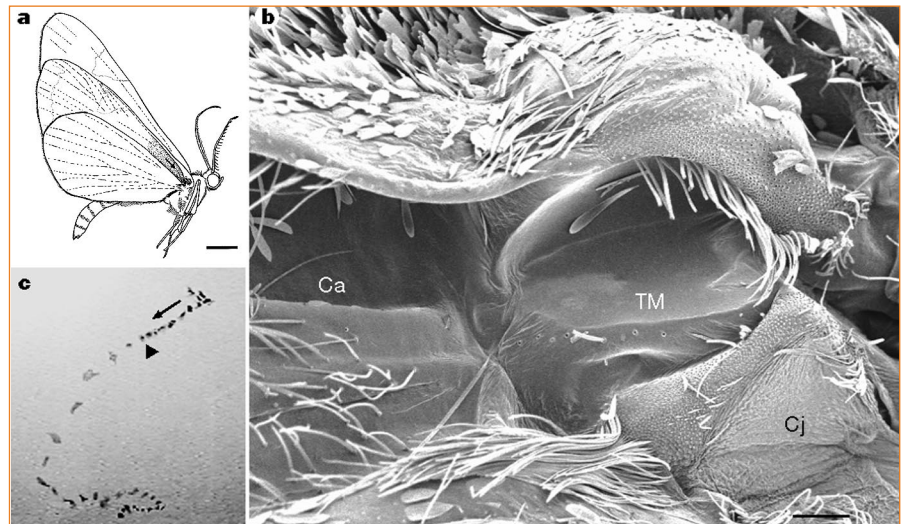


Figure 1 Hedyliid butterflies have ultrasound-sensitive ears that allow them to evade bats in flight. **a**, Lateral view of *M. heliconiaria*, showing the location of the right ear. An arrow points down the canal to the tympanic cavity where the tympanic membrane (not shown) resides. Scale bar, ~3 mm. **b**, Scanning electron micrograph of the right tympanic cavity showing the tympanic membrane (TM), ear canal (Ca) and an accessory membrane, the conjunctivum (Cj). Scale bar, 120 μ m. **c**, Consecutive video images (41 frames at 30 frames per s; 13 before and 28 after the stimulus onset) of a free-flying *M. heliconiaria* responding to a short (~250 ms), high-frequency (25 kHz), high-intensity (> 100 dB) sound. Arrow, flight direction; arrowhead, stimulus onset.

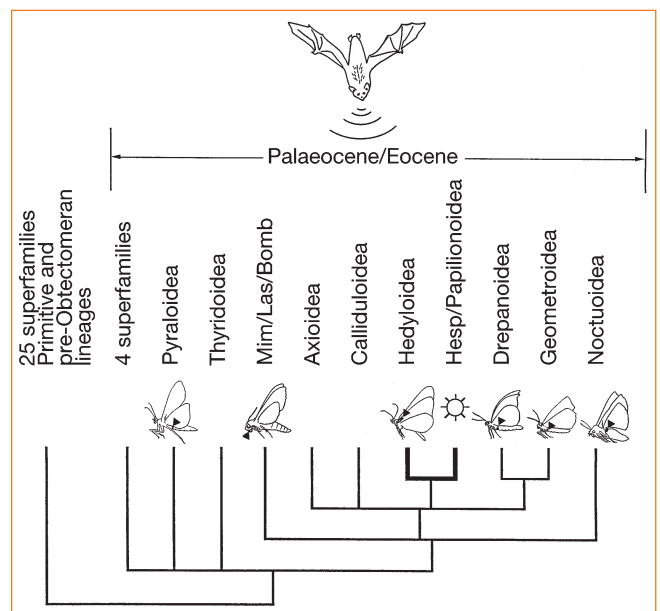
widespread throughout the Hedyloidea (about 40 species), and are equally developed in both sexes.

When stimulated with an intense ultrasonic stimulus, hedyliids perform one of several flight manoeuvres (Fig. 1c), including steep dives or climbs, upward or downward loops, spirals and horizontal sweeps. These responses, shown through ablation studies to be governed by the tympanal ears, are unpredictable in their direction, and are characterized by short latencies

(45.4 $\pm$ 9.35 ms) and substantial increases (400.1%) in flight speeds. These features are of selective advantage to an insect in a close-range encounter with a bat⁹.

The discovery of hearing in the Hedyloidea provides an insight into the evolution of lepidopteran ears and diurnality in butterflies. Mapping ultrasonic hearing onto a phylogeny of the Lepidoptera¹⁰ shows that it has evolved independently at least six times, all within the obtectomeran lineage (Fig. 2). Current fossil evidence indicates that the

Figure 2 The presence of hearing mapped onto a lepidopteran phylogeny¹⁰. Ultrasound-sensitive ears occur on a variety of body parts (arrowheads) reflecting multiple evolutionary origins. All ears, known to function primarily as bat-detectors, occur within the Obectomera. We propose that hearing in extant Lepidoptera originated as a response to selection pressures imposed by echolocating bats in the late Palaeocene, and that echolocating bats were a driving force for the large, species-rich butterfly groups Hesperoidea and Papilionoidea (Sun symbol) to move into the daytime. The butterfly superfamilies are highlighted in bold. Earlessness and nocturnalism are the plesiomorphic (ancestral) characteristics of the Lepidoptera.



principal obtectomeran groups originated around the late Palaeocene epoch (~60 million years ago)¹⁰, by which time microchiropteran bats had evolved echolocation¹¹. We conclude that predation by bats imposed a great selection pressure on the evolution of ears in Lepidoptera.

Butterflies are the largest and most diverse group of diurnal Lepidoptera, and the selection pressures generally proposed for their diurnality include various physiological and ecological factors, but not selection pressure by bats. Given the significant impact of bats on other obtectomeran taxa, we suggest that diurnality in non-hedylid butterflies was also an anti-bat strategy, promoted by selection for individuals that avoided bats by appearing during the day. The butterfly, in effect, was therefore 'invented' by the bat.

Is the earless condition of other (non-hedylid) butterflies primitive or secondarily derived? Consider the Vogel's organ, a forewing structure of unknown function that is distributed sporadically with varying degrees of development among certain Papilionoidea^{12,13}. Our comparative anatomical studies show that the hedylid ear and Vogel's organ are homologous structures. Given the current placement of the Hedy-

loidea as a sister-group to the Papilionoidea and Hesperoidea, it is possible that Vogel's organ is a degenerate 'bat detector'. Our discovery may help to bring to light the evolutionary origin of this group of butterflies.

Jayne E. Yack*, James H. Fullard†

*Department of Biology and College of Natural Sciences, Carleton University, Ottawa, Ontario K1S 5B6, Canada

†Department of Zoology and Erindale College, University of Toronto, Mississauga, Ontario L5L 1C6, Canada

e-mail: jyack@ccs.carleton.ca

1. Fullard, J. H. in *Comparative Hearing: Insects* (eds Hoy, R. R., Popper, A. N. & Fay, R. R.) 279–326 (Springer, New York, 1998).
2. Conner, W. E. *J. Exp. Biol.* **202**, 1711–1723 (1999).
3. Silberglied, R. E. in *The Biology of Butterflies* (eds Vane-Wright, R. I. & Ackery, P. R.) 207–223 (Academic, London, 1984).
4. Yack, J. E. & Fullard, J. H. *Ann. Entomol. Soc. Am.* **86**, 677–682 (1993).
5. Ribaric, D. & Gogala, M. *Acta Entomol. Slov.* **4**, 5–12 (1996).
6. Scoble, M. J. *Bull. Br. Mus. Nat. Hist. (Entomol.)* **53**, 251–286 (1986).
7. Scoble, M. J. & Aiello, A. *J. Nat. Hist.* **24**, 159–164 (1990).
8. Jacobs, K., Otte, B. & Lakes-Harlan, R. *J. Exp. Zool.* **283**, 270–285 (1999).
9. Hoy, R. R., Nolen, T. & Brodfuehrer, P. *J. Exp. Biol.* **146**, 287–306 (1989).
10. Kristensen, N. P. & Skalski, A. W. in *Handbook of Zoology, Arthropoda: Insecta, Part 35, Vol. 1, Lepidoptera: Moths and Butterflies* (ed. Kristensen, N. P.) 7–25 (de Gruyter, Berlin, 1998).
11. Novacek, M. J. *Nature* **315**, 140–141 (1985).
12. Vogel, R. *Z. Wiss. Zool.* **100**, 210–244 (1912).
13. Otero, L. D. *Bol. Entomol. Venez.* **5**, 128–138 (1990).

Alloys

Atomic structure of the quasicrystal $\text{Al}_{72}\text{Ni}_{20}\text{Co}_8$

Steinhardt *et al.*¹ reported experimental evidence in support of the coverage model^{2–4} by presenting a structure for the high-perfection decagonal quasicrystal $\text{Al}_{72}\text{Ni}_{20}\text{Co}_8$. The coverage model describes the decagonal quasicrystals by a single type of tile, a decagon, the basic 2-nm cluster of which is allowed to overlap so as to cover the surface. Although their Z-contrast image (also referred to as a high-angle annular dark-

field image) supports the coverage model, we find that the atomic structure they propose for the decagon has significant shortcomings which are inconsistent with our Z-contrast images.

The correct atomic structure of the decagon is critical to understanding how and why quasicrystals form. We have investigated the same high-perfection decagonal quasicrystal $\text{Al}_{72}\text{Ni}_{20}\text{Co}_8$ (provided by A. P. Tsai) by Z-contrast imaging, but at a higher spatial resolution (~0.13 nm) than Steinhardt *et al.*¹.

The main feature in the proposed structure of the decagon¹ is the extensive broken

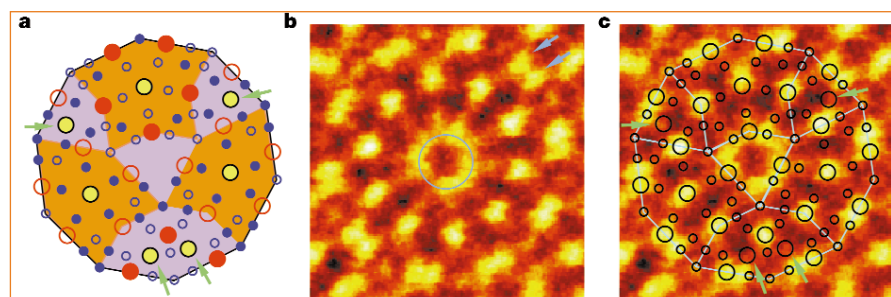


Figure 1 Comparison of the proposed "best-fit candidate model"¹ for the 2-nm cluster of the high-perfection decagonal quasicrystal $\text{Al}_{72}\text{Ni}_{20}\text{Co}_8$ with our higher resolution Z-contrast image. **a**, The proposed model contains broken decagonal symmetry for the entire cluster: see, for example, the four yellow columns indicated by green arrows. **b**, Z-contrast image of a 2-nm cluster of the same high-perfection decagonal quasicrystal $\text{Al}_{72}\text{Ni}_{20}\text{Co}_8$ along the ten-fold axis. The bright yellow features show the locations of highest intensity corresponding to high transition-metal (TM) occupancy. Al columns have lower intensity and are seen as red features: see, for example, the ten red spokes around the central ring. These spokes connect the central ring to the ring of ten TM columns (yellow) and show the basic ten-fold symmetry of the cluster. **c**, The model cluster of Steinhardt *et al.* superimposed on our Z-contrast image shows that their four proposed TM columns indicated by green arrows are not present.

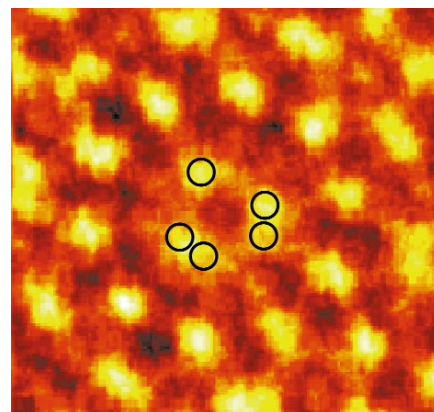


Figure 2 Cluster showing broken decagonal symmetry within the central ring, but there are still ten columns present, five of which have higher intensities (yellow), indicating high TM occupancy, whereas the other five show high Al occupancy (red). One of the five TM columns is a single column; the others form closely spaced column pairs, very similar to those in the outer 2-nm ring.

decagonal symmetry for the entire cluster (Fig. 1a) which enables the decagons to have identical subtiles, as suggested by the Gummelt coverage model, ensuring that the overlap rules required by this model are satisfied.

However, it is hard to see how a structure with such extensive broken symmetry can be energetically stable. Figure 1b shows a typical Z-contrast image of clusters that do not have strong broken symmetry. In a Z-contrast image, the intensity is directly correlated with the mean-square atomic number (Z), so that transition metal (TM) columns are seen with much higher intensity than Al columns.

The superimposition in Fig. 1c shows that the structure of the decagon proposed by Steinhardt *et al.* does not match our Z-contrast image in significant ways: the four proposed TM columns causing the broken symmetry are definitely absent from our image. In addition, our image reveals the presence of ten closely spaced TM column pairs in the outermost ring of the cluster, as indicated by double blue arrows in Fig. 1b. These are only single TM columns in the model of Steinhardt *et al.*

Turning to the central ring, our image clearly shows its underlying ten-fold symmetry. It is seen as a ring, with an intensity varying between that of an Al column (red) and a TM column (yellow). This is inconsistent with Steinhardt *et al.*'s triangular arrangement. This intensity pattern shows that there are ten closely spaced atomic columns around the central ring with a composition intermediate between that of an Al column and a TM column. However, there are many clusters with broken symmetry in the central ring.

Figure 2 is a typical Z-contrast image of such a cluster where the intensity in the central ring shows broken decagonal symmetry. The intensity distribution shows that

there are still ten closely spaced columns at the central ring, but now only five have higher intensities indicating high TM occupancy, whereas the intensity of the remainder is closer to that of an Al column. This indicates that the broken symmetry at the central ring is due to chemical ordering, and that there are ten sites, but with different TM and Al occupancies. This is also inconsistent with the triangular arrangement proposed by Steinhardt *et al.*

There are significant differences between the structure model proposed by Steinhardt *et al.* and our atomic-resolution Z-contrast image. Although these do not invalidate the coverage picture, they do prevent our understanding the formation of quasicrystals.

We believe that the closely spaced column pairs in the central and outermost rings that were not predicted by the structure model of Steinhardt *et al.* are the key to understanding the formation of decagonal quasicrystals. They not only show that the structures of the decagonal quasicrystals and their crystalline approximants are more similar than Steinhardt *et al.* suppose, they also highlight the critical differences. On this basis, we have proposed a growth mechanism⁵ that explains why these clusters prefer to overlap and follow the Gummelt coverage picture. Our growth model predicts that the overall structure will show ideal quasicrystal tiling, in the Gummelt coverage picture, when all clusters have strong chemical ordering in the central rings. If the clusters have no chemical ordering, the model predicts a random tiling. For real quasicrystals, their structure might be a mixture of both cases.

Yanfa Yan*, **Stephen J. Pennycook**

Solid State Division,
Oak Ridge National Laboratory,
Oak Ridge, Tennessee 37831, USA

*Present address: National Renewable Energy
Laboratory, Golden, Colorado 80401, USA

e-mail: yanfa_yan@nrel.gov

1. Steinhardt, P. J. *et al.* *Nature* **396**, 55–57 (1998); correction *Nature* **399**, 84 (1999).
2. Gummelt, P. *Geometriae Dedicata* **62**, 1–17 (1996).
3. Steinhardt, P. J. & Jeong, H.-C. *Nature* **382**, 433–435 (1996).
4. Jeong, H.-C. & Steinhardt, P. J. *Phys. Rev. B* **55**, 3520–3532 (1997).
5. Yan, Y. & Pennycook, S. J. *Mater. Sci. Eng.* (in the press).

Steinhardt *et al.* reply — The purpose of our Letter¹ was to present experimental support for the quasi-unit cell picture of quasicrystals. This model proposes that the atomic structure can be reduced to a single repeating cluster satisfying certain ‘overlap rules’ (sharing of atoms by neighbouring clusters). We proposed that the quasicrystalline phase of AlNiCo can be decomposed into a repeating decagonal atom cluster (20 Å in radius). Yan and Pennycook do not refute the quasi-unit cell concept — they also propose a repeating cluster obeying the same

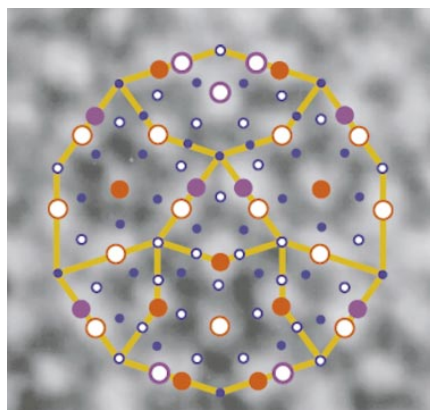


Figure 1 Improved decoration of the quasi-unit cell for AlNiCo compared to lattice image. Problematic TM sites in our earlier model¹ have been removed. The figure includes atoms added by overlap of neighbour clusters; these lead to the formation of neighbour TM column pairs, as seen near the centre. Large circles represent Ni (red) or Co (purple) and small circles represent Al. Solid circles represent $c=0$ and open circles represent $c=1/2$ along the periodic c -axis.

overlap rules. However, they propose a different atomic decoration for the repeating cluster that is ten-fold symmetric, whereas our decoration explicitly breaks ten-fold symmetry. This is important because our symmetry breaking corresponds precisely to the symmetry breaking of the overlap rules, and hence provides key evidence for the quasi-unit cell picture.

Yan and Pennycook's decoration is motivated by their impressive high-angle annular dark-field (HAADF) imaging, obtained with higher resolution than we had available. As they show, the imaging disagrees with the sites of four columns of transition metal (TM) atoms in our proposal (shown by arrows in their Fig. 1). However, we find that the problem can be resolved by a modest rearrangement of the previous decoration, switching 8 out of 100 atoms and retaining the broken ten-fold symmetry. The improved model in Fig. 1 has all the same qualitative properties as the original in ref. 1, matches the new HAADF (including Yan and Pennycook's Fig. 2.) and even more recent high-resolution transmission electron microscopy (HRTEM) imaging, and has a density and stoichiometry that fits measured values to better than 2 per cent.

As more data become available (for example, from X-ray diffraction), further small refinements to our current best-fit decoration may be required, but the ten-fold symmetry breaking should remain as an essential property. The broken symmetry is necessary to explain three established features of AlNiCo: the broken symmetry consistently observed in through-focus HRTEM imaging of the clusters²; the broken symmetry found within the central ring of most clusters in HAADF imaging, such as our Fig. 1 and Yan and Pennycook's Fig. 2 (ref. 2) (the very rare, more symmetric

rings, as shown in their Fig. 1, can be explained as defects; ref. 2 and M. Widom, personal communication); and the apparent quasiperiodic correlation in the broken symmetry direction on moving from cluster to cluster in HAADF images (see Fig. 1 of ref. 1), as is found for a configuration of overlapping decagons.

None of the features can be explained by symmetric clusters, even if chemical disorder is introduced to randomly break the ten-fold symmetry. M. Widom and co-workers (personal communication) have completed a total-energy-based prediction of the structure of AlNiCo, making no prior assumption about the existence of repeating 20-Å clusters. Yet decagonal clusters with broken ten-fold symmetry emerge as the lowest-energy configuration with nearly identical assignments of Al and TM positions, as in our improved model.

Paul J. Steinhardt*, **H.-C. Jeong†**,
K. Saitoh‡, **M. Tanaka‡**, **E. Abe§**,
A. P. Tsai§

*Department of Physics, Princeton University,
Princeton, New Jersey 08544, USA

†Department of Physics, Sejong University,
Kwangjin, Seoul 143-747, Korea

‡Research Institute for Scientific Measurements,
Tohoku University, 2-1-1 Katahira,
Aoba-ku, Sendai 980-8577, Japan

§National Research Institute for Metals,
1-2-1 Sengen, Tsukuba,
Ibaraki 305-0047, Japan

1. Steinhardt, P. J. *et al.* *Nature* **396**, 55–57 (1998); correction *Nature* **399**, 84 (1999).
2. Abe, E. *et al.* <http://xxx.lanl.gov/abs/cond-mat/9907160>

Biological rhythms

Circadian clocks limited by noise

Circadian rhythms, which provide internal daily periodicity, are used by a wide range of organisms to anticipate daily changes in the environment¹. It seems that these organisms generate circadian periodicity by similar biochemical networks within a single cell². A model based on the common features of these biochemical networks shows that a circadian network can oscillate reliably in the presence of stochastic biochemical noise and when cellular conditions are altered. We propose that the ability to resist such perturbations imposes strict constraints on the oscillation mechanisms underlying circadian periodicity *in vivo*.

There is evidence that clock networks share common features in a wide range of organisms, from cyanobacteria to mammals². For instance, all networks seem to include an interaction between two types of component (Fig. 1a): positive elements (or activators, such as KaiA in *Synechococcus*, Wc1-2 in *Neurospora*, Clc and Cyc in

Drosophila, and Clock and Bmal in mice) enhance the expression of negative elements (or repressors, such as KaiB and KaiC in *Synechococcus*, Frq in *Neurospora*, Tim and Per in *Drosophila*, and Tim and Per1,2 and 3 in mice).

The evidence indicates that the clock network is based predominantly on transcriptional regulation². Although the contribution of post-transcriptional regulation to the oscillation mechanism is not yet clear³, and it is too early to be sure that all the components and their interactions have been revealed in any single organism, we can already address the question of whether the common features of circadian networks are consequences of some underlying 'design principles'.

We can envisage many types of biochemical network that produce periodic oscillations^{4,5} and that can be entrained to a 24-hour period by an external periodic stimulus, such as light or temperature. But there are additional constraints: for example, the periods of all autonomous circadian clocks must remain relatively constant over a wide temperature range, a property known as temperature compensation¹.

The ability to function reliably in the presence of internal noise may impose a further constraint on the oscillation mechanism. Internal noise in the operation of biochemical networks results from the stochastic nature of reaction events⁶⁻⁸, and is particularly important when there are few molecules in the system, as is often the case in a cell⁹. These effects have sometimes been overlooked in previous analyses of circadian rhythms. For example, in a previously studied model that depends on a time-delayed negative feedback, reliable oscillations were found when reaction kinetics were approximated by continuous differential equations¹⁰. However, when the discrete nature of reaction events is taken into account, the oscillations persist but with periods and amplitudes that fluctuate widely in time (Fig. 1c). Noise resistance should therefore be considered in any postulated molecular mechanism of circadian rhythms.

We can illustrate this point with an example, based on existing data, of a class of biochemical networks that sustain reliable oscillations even in the presence of internal noise. In this class of biochemical network, the interactions between positive and negative regulatory elements lead to a hysteresis in the dependence of protein expression rates on the concentration of the negative element (see Fig. 1a,b). It is possible to formulate different versions of the model based on both transcriptional and post-transcriptional regulation (Fig. 1a).

The temporal evolution of the different components of the network is shown in Fig. 1d,e. We used a Monte Carlo algorithm, in which molecules participate in stochastic

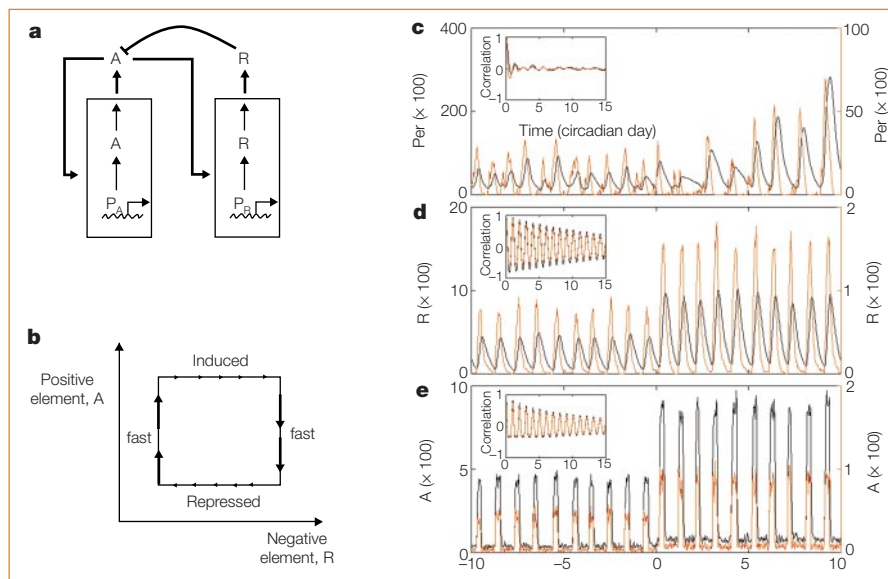


Figure 1 The hysteresis-based oscillation mechanism. **a**, A positive element, A, increases its own expression and that of a negative element, R. Strong binding of R to A inhibits A activity and so represses the expression of both elements by binding to the promoters P_A and P_R . **b**, The autoactivation of A results in a hysteresis: a bistable dependence of A concentration on R. Slow kinetics of R then leads to oscillations, which can be described as successive transitions between 'induced' and 'repressed' states. **c-e**, Monte Carlo simulations of models based on time delay (**c**) and hysteresis (**d,e**) for circadian oscillations. The time between two successive events is distributed exponentially around the usual mass-action reaction rate⁶. At time $t = 0$, all transcription rates were doubled. Black curves indicate proteins (left axis); orange curves, mRNA (right axis). Inset, temporal autocorrelation functions. Reliable oscillations result in temporal autocorrelations that persist for many periods. **c**, Model exhibiting reliable oscillation in the continuous limit⁴; we use the same parameter values as in Fig. 2 of ref. 4 (assuming that binding to the DNA promoter is diffusion limited, and that 1 nM corresponds to 1,000 molecules per cell). **d, e**, Simulation of the hysteresis-based oscillator. A is assumed to be a transcriptional activator. Further details are available from the authors.

reaction events⁶. The time between two successive occurrences of a specific reaction is exponentially distributed around its mean, which is given by the usual mass-action reaction rate (the same algorithm was used to simulate the noise effects for the time-delay oscillator in Fig. 1c). For a wide range of parameter values, the oscillations are reliable, with long time correlations (see insets in Fig. 1d,e) and small variations in period length, even with relatively small numbers of molecules. Reducing the transcription rate of the system depicted in Fig. 1d,e leads to oscillations in which average levels of messenger RNA are as low as 10 molecules per cell. In these systems, the standard deviation of the period remains less than 10%.

A further consideration is whether the circadian circuit can operate reliably within the cellular context. Global changes in transcription and translation rates may arise from variations in nutrition, growth conditions or temperature, and may affect the period of transcription or translation-based oscillators (Fig. 1c). In the hysteresis-based model, global transcription or translation rates have only small effects on the period, but changes in these rates alter the amplitude. The ability to maintain constant circadian periodicity despite global changes in the state of the cell¹¹ is probably necessary for the circadian clock to be successfully embedded within the cell.

It is not clear whether this hysteresis-based network is the mechanism underlying circadian oscillations. For instance, the regulation of positive and negative elements in the *Drosophila* clock might be more complex than this¹². Oscillation mechanisms differ in their sensitivity to internal noise and changes in cellular conditions, however, and the ability to resist such uncertainties was probably one of the decisive factors in the evolution of circadian clocks and should be reflected in the underlying oscillation mechanism. Further studies of noise resistance may therefore help to uncover the design principles underlying circadian clocks.

Naama Barkai, Stanislas Leibler

Departments of Physics and Molecular Biology,
Princeton University, Princeton,
New Jersey 08544, USA
e-mail: leibler@princeton.edu

1. Edmund, L. N. *Cellular and Molecular Basis of Biological Clocks* (Springer, New York, 1988).
2. Dunlap, J. *Cell* **96**, 271–290 (1999).
3. Suri, V., Lanjuin, A. & Rosbash, M. *EMBO J.* **18**, 675–686 (1999).
4. Goldbeter, A. *Biochemical Oscillations and Cellular Rhythms* (Cambridge Univ. Press, 1995).
5. Murray, J. D. *Mathematical Biology* (Springer, Berlin, 1994).
6. Gillespie, D. T. *J. Phys. Chem.* **81**, 2340–2361 (1977).
7. Ko, M. S. *Bioessays* **14**, 341 (1992).
8. McAdams, H. N. & Arkin, A. *Trends Genet.* **15**, 65–69 (1999).
9. Wodicka, J. et al. *Nature Biotechnol.* **15**, 1359–1367 (1997).
10. Leloup, J. C. & Goldbeter, A. *J. Biol. Rhythms* **13**, 70–87 (1998).
11. Kondo, T. et al. *Science* **275**, 224–227 (1997).
12. Glossop, N. R. J., Lyons, L. C. & Hardin, P. E. *Science* **286**, 766–768 (1999).

Decoherence of quantum superpositions through coupling to engineered reservoirs

C. J. Myatt*, B. E. King*, Q. A. Turchette, C. A. Sackett, D. Kielpinski, W. M. Itano, C. Monroe & D. J. Wineland

National Institute of Standards and Technology, Div. 847.10, 325 Broadway, Boulder, Colorado 80303, USA

The theory of quantum mechanics applies to closed systems. In such ideal situations, a single atom can, for example, exist simultaneously in a superposition of two different spatial locations. In contrast, real systems always interact with their environment, with the consequence that macroscopic quantum superpositions (as illustrated by the ‘Schrödinger’s cat’ thought-experiment) are not observed. Moreover, macroscopic superpositions decay so quickly that even the dynamics of decoherence cannot be observed. However, mesoscopic systems offer the possibility of observing the decoherence of such quantum superpositions. Here we present measurements of the decoherence of superposed motional states of a single trapped atom. Decoherence is induced by coupling the atom to engineered reservoirs, in which the coupling and state of the environment are controllable. We perform three experiments, finding that the decoherence rate scales with the square of a quantity describing the amplitude of the superposition state.

One of the fundamental properties of quantum mechanics is the principle of superposition, a principle whose introduction was considered a “drastic” measure by Dirac¹. The fact that quantum superpositions do not exist in the macroscopic world hinders our intuition and leads to the apparently strange behaviour dictated by quantum mechanics. A famous example of this was posed by Schrödinger in 1935 (ref. 2) who pointed out that quantum mechanics would predict bizarre situations such as a cat being simultaneously dead and alive. The existence of superpositions prescribed by quantum mechanics is valid for systems that are closed, that is, free from external influences. In contrast, real systems always couple to these external influences, the environment, which is typically composed of an extremely large number of degrees of freedom. Lack of knowledge about the environment is expressed by averaging (mathematically tracing) over the possible states of the environmental degrees of freedom. This leads to an evolution of the density matrix of the system, in which the quantum superpositions are continuously reduced to classical probability distributions, a process generally known as decoherence (see, for example, refs 3–5). One approach to describing decoherence is to treat the environment as a reservoir of quantum oscillators, each of which interacts with the quantum system in question. An example of such a reservoir–system interaction is the ensemble of empty electromagnetic field modes, each represented by a quantized harmonic oscillator, interacting with an atom in order to induce spontaneous emission. As a quantum superposition is made larger, decoherence tends to act more quickly. For truly macroscopic superpositions, such as that of ‘Schrödinger’s cat’, decoherence occurs on such a short timescale that it is almost impossible to observe quantum coherences. However, mesoscopic systems present the possibility of studying, in a controlled way, the process of decoherence and the transition from quantum to classical behaviour.

In the past few years, techniques have been realized to generate mesoscopic superpositions, also called ‘Schrödinger cats’, of motional states of trapped ions⁶ and of photon states in the context of cavity QED (ref. 7), where decoherence through coupling to

ambient reservoirs and the sensitivity of the rate of decoherence to the size of cat were observed. Here we extend the investigations beyond the ambient reservoirs and ‘engineer’ the state of the reservoir, as well as the form of the system–reservoir coupling. One way this can be achieved for a system of trapped ions is by applying noisy potentials to the trap electrodes, simulating a hot resistor (reservoir) connected to the trap electrodes, with controllable temperature and spectrum. For a range of two-component superposition states, we demonstrate the expected exponential dependence of the decoherence rate on the separation of the components in Hilbert space. We also present the first, to our knowledge, study of decoherence into an engineered quantum reservoir, using laser cooling techniques to generate an effectively zero-temperature bath^{8,9}.

Theoretical predictions

Decoherence of specific mesoscopic quantum superpositions, with a variety of couplings to a reservoir, has been investigated extensively in theory^{3–5,8,10–12}. The model in these studies is a system harmonic oscillator coupled to a bath of environment quantum oscillators. (These and other sources of decoherence in the context of trapped-ion experiments have been more recently discussed theoretically in refs 13–16.) As an illustration, we consider the system oscillator to be in a superposition of coherent states. A coherent state¹⁷ of a harmonic oscillator is a gaussian wavepacket which oscillates back and forth while retaining its shape. In quantum mechanics, a coherent state is represented by a state vector $|\alpha\rangle$, where $\alpha = |\alpha|e^{i\theta}$ is a complex number whose magnitude $|\alpha|$ is a dimensionless amplitude of the wavepacket’s motion and whose phase θ is the phase of the oscillation at some initial time $t = 0$ (the phases of all subsequent coherent manipulations are set relative to this initial phase). Coherent states are analogous to classical trajectories of a harmonic oscillator, approximated by a marble rolling back and forth in a bowl. A superposition of coherent states, $|\psi\rangle = N(|\alpha_1\rangle + |\alpha_2\rangle)$ where N is a normalization factor, can be visualized as a marble rolling in a superposition of two trajectories.

We consider the system oscillator to couple to the reservoir through an interaction proportional to the product of the amplitude of motion of the system oscillator and the amplitude of

* Present addresses: Research Electro-Optics, 1855 South 57th Court, Boulder, Colorado 80301, USA (C.J.M.); NIST, Atomic Physics Division (842), 100 Bureau Drive, Stop 8424, Gaithersburg, Maryland 20899-8424, USA (B.E.K.).

fluctuations of the reservoir. For brevity, we call this an amplitude reservoir. In the classical analogy, a hot amplitude reservoir behaves as if the bowl is subject to random displacements of its centre, resulting in a random force on the marble. For a superposition of coherent states coupled to such a reservoir, a simple scaling law may be stated: the rate of decoherence (here a dephasing between the $|\alpha_1\rangle$ and $|\alpha_2\rangle$ components of $|\psi\rangle$) scales as the square of the separation of the wave packets, $|\alpha_1 - \alpha_2|^2$. In an idealized case where, first, the superposition is created, then the amplitude reservoir is coupled to the system for a time t , and then the coupling is turned off, the remaining coherence between the two wave packets is⁴:

$$C(t) = \exp[-|\alpha_1 - \alpha_2|^2 \xi t] \quad (1)$$

Here ξ is a coupling constant between the reservoir and the system. The larger the size ($|\alpha_1 - \alpha_2|$) of the superposition, the faster the decoherence.

Another basis of quantum states for the harmonic oscillator is the energy eigenstates, also known as Fock or number states. The Fock state $|n\rangle$ has energy $\hbar\omega(n + 1/2)$ and represents a state of n units of quantized vibration, where $n \geq 0$ is an integer. Fock states have no classical analogue, as they are delocalized in position and uniformly distributed in phase. A superposition of two Fock states $|\psi\rangle = (|n_1\rangle + |n_2\rangle)/\sqrt{2}$ loses coherence when the modes of the reservoir couple linearly to the energy of the oscillator, which is equivalent to averaging over a gaussian distribution of phase shifts of the oscillator. We denote this case a phase reservoir. The coherence between the two Fock states decays at a rate that scales as the square of the difference between the Fock indices, $|n_1 - n_2|^2$, given by⁴:

$$C(t) = \exp[-|n_1 - n_2|^2 \kappa t] \quad (2)$$

Here κ is a coupling constant.

Trapped ions

In the experiments described here, a linear Paul trap, similar to the one described in ref. 18, confines single ${}^9\text{Be}^+$ atomic ions in a harmonic potential, for which we isolate the axial motion at frequency $\omega = 2\pi \times 11.3$ MHz. Within the ion's electronic ground-state hyperfine manifold we restrict our attention to two states, the $|F = 2, m_F = -2\rangle$ state, which we label $|\downarrow\rangle$, and the $|F = 1, m_F = -1\rangle$ state, which we label $|\uparrow\rangle$, separated in energy by

$\hbar\omega_0$, where $\omega_0 \approx 2\pi \times 1.25$ GHz, and where F and m_F are the quantum numbers associated with the total angular momentum of the atomic state. The ion is cooled to the $n = 0$ ground state of motion, denoted $|0\rangle$, and optically pumped to the $|\downarrow\rangle$ state with resolved-sideband stimulated Raman cooling¹⁹. Thus, the initial state for all the experiments is $|\downarrow\rangle|0\rangle$.

We drive coherent stimulated Raman transitions with a pair of laser beams detuned approximately 12 GHz from the atomic resonance near 958 THz ($\lambda = 313$ nm). We use three types of Raman transitions, determined by the beam geometry and difference frequency of the two beams: (1) Motion-independent spin-flip transitions ($|\downarrow\rangle|n\rangle \leftrightarrow |\uparrow\rangle|n\rangle$). Here, the Raman beams are co-propagating and the difference frequency is set to ω_0 . (2) Sideband transitions ($|\downarrow\rangle|n\rangle \leftrightarrow |\uparrow\rangle|n + \Delta n\rangle$). Here, the beams are orientated with their difference wavevector pointing along the trap axis and their difference frequency set to a motional sideband at $\omega_0 + \omega\Delta n$. (3) Motional displacement transitions. Here, the beams are orientated with their difference wavevector pointing along the trap axis and their difference frequency set to the trap frequency ω . This approximates the harmonic-oscillator displacement operator $D(\alpha)$, where the operator is defined^{4,5} by the relation $D(\alpha)|0\rangle = |\alpha\rangle$. The displacement $|\alpha|$ is proportional to the duration of the laser pulse, and θ is set by the phase of the applied laser field^{6,13}. In general, α depends on the internal state of the ion.

We can efficiently detect the $|\downarrow\rangle$ internal state of the ion by applying circularly polarized laser light resonant with the transition $|\downarrow\rangle \leftrightarrow |e\rangle$, where $|e\rangle$ is a short-lived excited electronic state that usually decays back to $|\downarrow\rangle$ by emitting a photon¹⁹. In contrast, the transition $|\uparrow\rangle \leftrightarrow |e\rangle$ is out of resonance, and an ion in the $|\uparrow\rangle$ state scatters negligible light.

High-temperature amplitude reservoir

The motion of a trapped ion couples to uniform electric fields $\mathbf{E}$ through the potential $U = -q\mathbf{x} \cdot \mathbf{E}$, where $\mathbf{x}$ is the displacement of the ion from its equilibrium position (proportional to the amplitude of motion) and q is the charge of the ion. This coupling is independent of the ion's internal state. Our engineered amplitude reservoir consists of random uniform electric fields applied along the axis of the trap, oscillating near the ion's axial-motion frequency ω . We generate axial fields in the trap by applying voltages to one of the trap electrodes. A commercial function generator produces

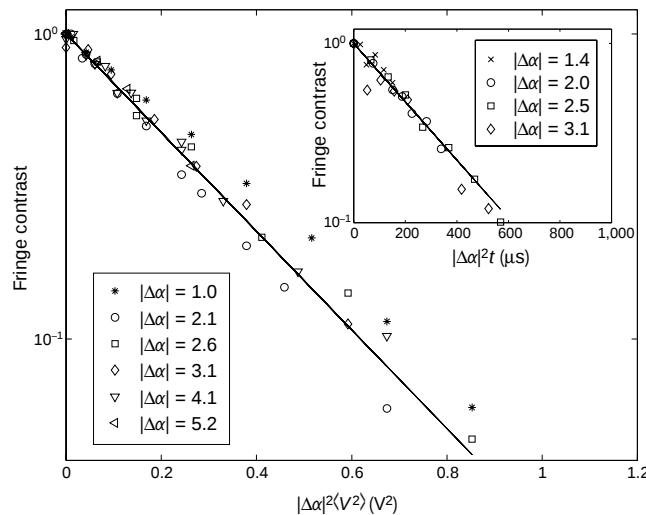


Figure 1 The decoherence of ‘Schrödinger-cat’ states coupled to an amplitude reservoir. In the main figure, each point is the measured contrast of the interference fringes after noisy potentials were applied to the trap electrodes. The fringe contrast at $\langle V^2 \rangle = 0$ is scaled to unity in order to make comparisons between different values of $|\Delta\alpha|$. The size of the superposition, $|\Delta\alpha|$, varies linearly with the pulse time for Raman transition type (3).

The applied mean-squared voltage $\langle V^2 \rangle$ is scaled by $|\Delta\alpha|^2$. The solid line is a fit to an exponential. Inset, fringe contrast versus time of interaction with ambient fields is plotted. Again, the fringe contrast is scaled to unity at $t = 0$ for comparison between different values $|\Delta\alpha|$. The solid line is a fit to an exponential.

pseudo-random voltages which are applied through a band-pass filter centred near ω , defining the frequency spectrum of the reservoir.

In all the experiments reported here, we measure the coherence of the quantum superpositions with single-atom interferometry, analogous to that used in our previous work⁶. For example, to observe the effects of the amplitude reservoir, the motional state of the ion is split into a superposition of two components, each associated with a different internal state of the ion, forming a state like that of the Schrödinger cat⁶. The superposition is then coupled to the reservoir, and finally the perturbed superposition is recombined by reversing the steps which initially created it. We repeat the experiment many times, measuring the internal state of the ion as a function of the relative phase of the creation and reversal steps, and the contrast of the resulting interference fringes characterizes the amount of coherence remaining after coupling to the reservoir.

In more detail, we first form a cat state of the form:

$$|\psi_c\rangle = (|\downarrow\rangle|\alpha_1\rangle + |\uparrow\rangle|\alpha_1\rangle)/\sqrt{2} \quad (3)$$

This is created by driving a Raman transition (type (1)) to generate an equal spin superposition, $|\downarrow\rangle|0\rangle \rightarrow (|\downarrow\rangle + |\uparrow\rangle)|0\rangle/\sqrt{2}$, followed by a Raman transition (type (3)), with laser polarizations set such that $\alpha_1 = -\alpha_i/2$ in equation (3).

A uniform electric field oscillating near the trap frequency ω (applied in the experiment for 3 μ s) results in the displacement operator $D(\beta)$ acting equally on both $|\downarrow\rangle$ and $|\uparrow\rangle$, giving:

$$|\psi_c\rangle \rightarrow |\psi'_c\rangle = (|\downarrow\rangle|\beta + \alpha_1\rangle + e^{i\phi_m}|\uparrow\rangle|\beta + \alpha_1\rangle)/\sqrt{2} \quad (4)$$

Here $\phi_m = \text{Im}\beta\Delta\alpha^*$ and $\Delta\alpha = \alpha_i - \alpha_1$. We probe the coherence by reversing the steps taken to generate the cat state. We first reverse the motional Raman transition (type (3)), resulting in the state

$$|\psi'_c\rangle \rightarrow |\psi''\rangle = (|\downarrow\rangle + e^{2i\phi_m}|\uparrow\rangle)|\beta\rangle/\sqrt{2} \quad (5)$$

A final pulse on the motion-independent spin-flip transition (1), with phase δ relative to the first pulse on transition (1), leads to interference fringes with a residual phase shift $2\phi_m$. Averaging over the gaussian random variable β , the probability of finding

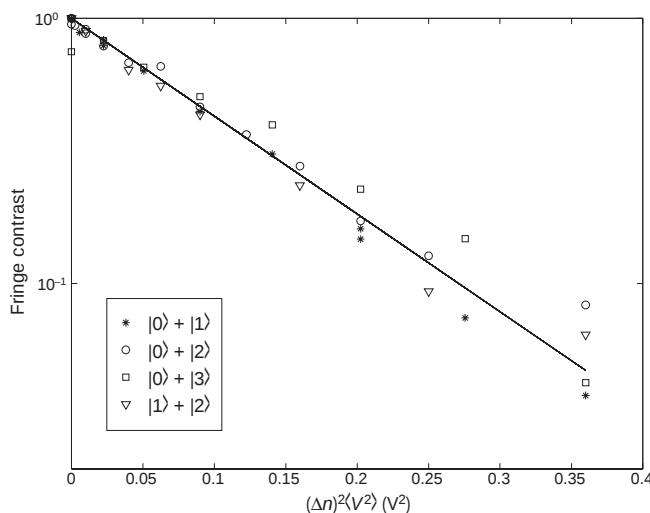


Figure 2 Decoherence of superpositions of Fock states coupled to the phase reservoir. The data points are the measured fringe contrast. The fringe contrast is normalized to unity at $\langle V^2 \rangle = 0$. The mean squared voltage applied to the trap electrodes is scaled by the squared size of the superposition $|\Delta n|^2$. The solid line is a fit to an exponential.

the ion in the $|\downarrow\rangle$ state is²¹:

$$P_{\downarrow} = \frac{1}{2}(1 - e^{-2|\Delta\alpha|^2\sigma^2}\cos\delta) \quad (6)$$

Interference fringes are generated by recording $P_{\downarrow}$ while sweeping δ . The variance σ^2 of β is proportional to the mean-squared voltage noise $\langle V^2 \rangle$ (proportional to the temperature of the simulated resistor). A plot of the interference-fringe contrast as a function of the applied mean-squared voltage, scaled by the squared 'size' of the cat state $|\Delta\alpha|^2$, is shown in Fig. 1. Decay curves were recorded for a variety of superposition sizes $|\Delta\alpha|$, and all the data agree with a single exponential.

In addition to the engineered reservoir of the applied voltage noise, the ion also interacts with ambient fluctuating electric fields, which we expect to have the character of an amplitude reservoir. To examine this 'natural' decoherence, we ran the experiment outlined above without any applied voltage noise, and with a variable time t between the creation of the cat state and the recombination. The fringe visibility as a function of $|\Delta\alpha|^2 t$ is shown in the inset to Fig. 1. The decay curves are normalized to unity at $t = 0$. The decay of the fringe visibility is exponential, and the decay constant $\gamma \approx 6.7 \times 10^{-3} \mu\text{s}^{-1}$ is consistent with the measured heating rate¹³ of $\gamma \approx 5.9 \times 10^{-3} \mu\text{s}^{-1}$ for this apparatus. The effects of this ambient reservoir were negligible during the time (3 μ s) that the engineered amplitude reservoir was coupled to the ion.

High-temperature phase reservoir

A phase reservoir coupled to the ion is simulated by random variations in the trap frequency ω , changing the phase of the ion oscillation without changing its energy. We realize this coupling experimentally by modulating the trap frequency. A random voltage noise source is passed through a low-pass filter network with a cut-off frequency well below ω to maintain adiabaticity. The fluctuations in potential are applied symmetrically to the trap electrodes so as to produce linear field gradients and negligible uniform fields. This in turn perturbs the trap frequency, by $\delta\omega(t)$. When integrated over the time (20 μ s) of the applied noise, the ion's motion is phase-shifted by $\phi = \int \delta\omega(t)dt$. This technique yields a gaussian-distributed ensemble of phase shifts with variance σ^2 proportional to the applied mean-squared voltage noise $\langle V^2 \rangle$.

Motional decoherence caused by a phase reservoir is clearly illustrated with a superposition of two Fock states. We generate superpositions of Fock states of the form $|\psi\rangle = |s\rangle(|n\rangle + |n'\rangle)/\sqrt{2}$, where $s = \downarrow$ or $\uparrow$, with pulses on the Raman motional sidebands (case (2) above) as in ref. 20. The trap frequency is then perturbed by the

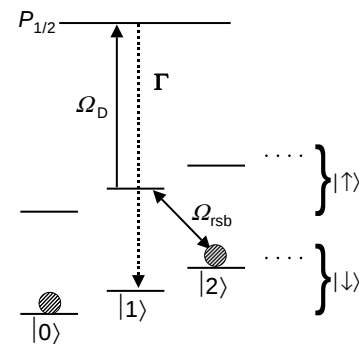


Figure 3 Implementation of an engineered zero-temperature reservoir. The states $|\downarrow\rangle|n\rangle$ and $|\uparrow\rangle|n-1\rangle$ are coupled by driving Raman motion-sensitive transitions (case (2) in the text). The state $|\uparrow\rangle$ is coupled to the environment by applying a weak optical pumping beam. The circles represent the superposition generated before applying this zero-temperature reservoir. The arrows show how the population in the $|\downarrow\rangle|2\rangle$ state is driven to the $|\uparrow\rangle|1\rangle$ state, and subsequently to the $|\downarrow\rangle|1\rangle$ state, through spontaneous Raman scattering.

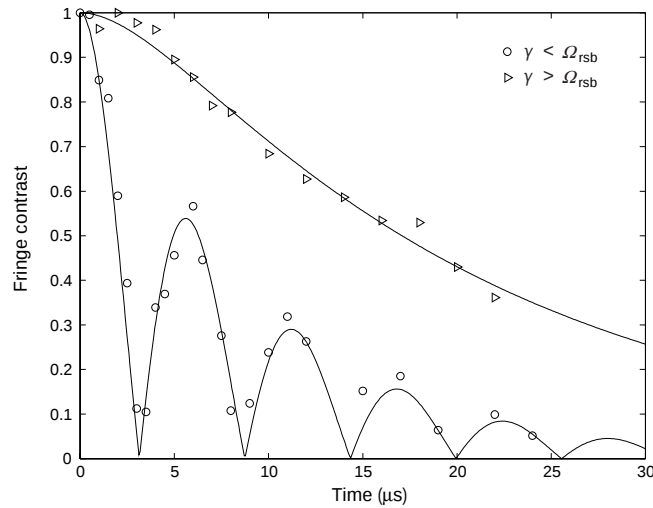


Figure 4 Decoherence of a Fock state superposition into the engineered zero-temperature reservoir. Fringe contrast is plotted as a function of the time the system is

applied random potentials, and the Fock states of the superposition acquire a relative phase factor $e^{i\psi\Delta n}$, where $\Delta n = n - n'$. The steps that created the superposition are then reversed, with a relative phase difference δ between the creation and reversal pulses, leading to a probability of detecting the ion in the $|\downarrow\rangle$ state²¹:

$$P_{\downarrow} = \frac{1}{2} [1 + e^{-|\Delta n|^2 \sigma^2 / 2} \cos \delta] \quad (7)$$

Interference fringes are recorded by varying δ as in the Schrödinger-cat interferometer. The fringe contrast is plotted as a function of $|\Delta n|^2 \langle V^2 \rangle$ in Fig. 2. As with the Schrödinger-cat states and amplitude reservoir, the data were fitted by a single exponential in $|\Delta n|^2 \langle V^2 \rangle$.

Zero-temperature reservoir

A third type of engineered reservoir requires a quantum mechanical description. This is a bath of laser cooling light plus optical spontaneous emission, an engineered (nearly) zero-temperature reservoir following the suggestion of Poyatos *et al.*⁸. Our implementation, shown in Fig. 3, is essentially a continuous Raman cooling technique. A pair of Raman beams (case (2)), tuned to the first red sideband, couples the states $|\downarrow\rangle|n\rangle \leftrightarrow |\uparrow\rangle|n-1\rangle$. Concurrently, an optical pumping beam causes spontaneous Raman transitions from $|\uparrow\rangle$ to $|\downarrow\rangle$ through an unstable excited state $|e\rangle$, which decays at rate Γ . The Raman coupling strength is characterized by the Rabi frequency Ω_{rsb} , a function of the intensity and detuning of the Raman beams¹³. If the Rabi frequency of the optical pumping beam is Ω_D , then we can define an effective damping rate for the $|\uparrow\rangle$ state of $\gamma = \Omega_D^2 / \Gamma$, valid for our case of $\Omega_D \ll \Gamma$. From the diagram in Fig. 3 we see that all populations are driven towards the state $|\downarrow\rangle|0\rangle$, the defining property of a zero-temperature reservoir. By varying the strength of the Raman and optical pumping couplings, we can control the reservoir parameters.

In the experiment, we examine the time evolution of the coherence of the Fock state superposition $\psi = |\downarrow\rangle(|0\rangle + |2\rangle)/\sqrt{2}$ for varying lengths of reservoir-interaction time. The interferometry is the same as in the study of the phase reservoir, where the Fock superposition is created, coupled to the reservoir, recombined, and probed, generating interference fringes. The data are shown in Fig. 4. Each data point represents the contrast of the fringes after the system interacts with the reservoir. We show two cases, $\gamma < \Omega_{\text{rsb}}$ and $\gamma > \Omega_{\text{rsb}}$. In the former case, the coherence between the $|0\rangle$ and $|2\rangle$ state disappears and reappears over time, with an overall decay of the fringe contrast. The underlying effect is population transfer back and forth (Rabi flopping) between the states $|\downarrow\rangle|2\rangle$ and $|\uparrow\rangle|1\rangle$ with a coupling of the $|\uparrow\rangle|1\rangle$ state to the outside environment through

coupled to the zero temperature reservoir. The only difference in the two cases shown was the intensity of the optical pumping beam (see Fig. 3).

spontaneous Raman scattering. In effect, we have restricted the size of the environment (here the manifold of $|\uparrow\rangle|n\rangle$ states, weakly coupled to the outside environment) to an extent where we can reverse the effects of decoherence (of the $\psi = |\downarrow\rangle(|0\rangle + |2\rangle)/\sqrt{2}$ state) in a way similar to that proposed in ref. 22. This is also a striking example of non-exponential decay²³ in a context that is investigated in ref. 24. For the case $\gamma > \Omega_{\text{rsb}}$, the fringe contrast decreases monotonically to zero. Even in the case of monotonic decay, a deviation from exponential is observed, a manifestation of the quantum Zeno effect^{24,25}.

Although the data with $\gamma < \Omega_{\text{rsb}}$ illustrate how coherence transferred to the environment can be recovered, an alternative explanation would say that by transferring the $|\downarrow\rangle|2\rangle$ component of the superposition to the $|\uparrow\rangle|1\rangle$ state, we gain ‘which-path’ information in our interferometer—the paths being the $|\downarrow\rangle|0\rangle$ and $|\downarrow\rangle|2\rangle$ parts of the superposition. The oscillation in which-path information is analogous to that illustrated by other experiments^{26,27}.

Conclusions

The decoherence caused by the engineered high-temperature reservoirs described above can be explained by ensemble-averaging over random classical fields applied to the ion^{21,28}. From previous experiments²⁰, we know that we can undo the effects of this decoherence by applying, in each experiment, a pulse of radiation that reverses the ‘random’ displacement. Similarly, the experiments here could also be carried out by coupling a hot resistor (with appropriate spectral filtering) between the trap electrodes (our case would correspond to a limit where the temperature $T \rightarrow \infty$ and the damping resistance $R \rightarrow 0$) (ref. 10). However, even in this case we could, subject to both practical and fundamental measurement uncertainties, record the voltages applied to the electrodes and reverse the effects of the random noise in each experiment. If we choose to ignore any knowledge of the electrical potentials applied to the trap electrodes, we can account for the observations just as well by considering the ion to be coupled to a large number of quantum oscillators, forming a heat bath. In the latter case, the state of the ion is entangled with that of the environment oscillators. After tracing over the environment variables, we are left with a reduced system, involving only the ion. The behaviour is the same as that obtained in the former case, in which the decoherence is caused by a deliberately applied external potential, but the environment is not considered to be a dynamical system itself^{3,4}. Loosely speaking, the effect of an environment oscillator in the latter case is replaced by that of a single Fourier component of the electrical potential in the former case. Therefore, in the high-temperature limit simulated

by the first two experiments, one need not consider the entanglement with the environment because the environment noise can be sensed (classically) and its effects reversed. This is in contrast to the decay of ion motion into a zero-temperature reservoir described above, similar to that seen in cavity-QED experiments⁷. In this case, after the quantum system couples to the environment through spontaneous emission, a measurement of the environment is not sufficient to reverse the effects of decoherence.

The methods of engineering reservoirs that are presented here begin to broaden the field of experimental investigations of decoherence. With control over the reservoir parameters combined with non-classical motional states of trapped ions, detailed comparisons between theory and experiment are possible. Here we have simulated the decoherence caused by coupling a charged atom to a hot resistor (reservoir) by applying noisy voltages to the ion-trap electrodes. The cases considered demonstrate a quadratic dependence of the rate of decoherence on the size of the superpositions, demonstrating the difficulty in generating truly macroscopic superpositions, such as that of ‘Schrödinger’s cat’. As a practical matter, these ‘high-temperature’ sources of noise are important because they currently limit the performance of a trapped-ion quantum computer¹³. We have also simulated a zero-temperature reservoir by using laser cooling to damp the ion motion. Extensions of the technique used to generate this zero-temperature bath should permit some interesting system–bath interactions that would be difficult to realize in any other way. One possibility is generating a squeezed reservoir, where all initial states asymptotically relax to a squeezed state of motion⁸. Other couplings can be tailored to relax the system into a ‘Schrödinger-cat’ state^{29,30}. □

Received 24 September; accepted 30 November 1999.

1. Dirac, P. A. M. *The Principles of Quantum Mechanics* (Oxford Univ. Press, Clarendon, 1958).
2. Schrödinger, E. Die gegenwärtige situation in der quantenmechanik. *Naturwissenschaften* **23**, 807–812, 823–828, 844–849 (1935).
3. Zurek, W. H. Decoherence and the transition from quantum to classical. *Phys. Today* **44**, 36–44 (1991).
4. Walls, D. F. & Milburn, G. J. *Quantum Optics* (Springer, Berlin, 1994).
5. Bužek, V. & Knight, P. L. Quantum interference, superposition states of light, and nonclassical effects. *Prog. Opt.* **XXXIV**, 1–158 (1995).
6. Monroe, C., Meekhof, D. M., King, B. E. & Wineland, D. J. A “Schrödinger cat” superposition state of an atom. *Science* **272**, 1131–1136 (1996).
7. Brune, M. *et al.* Observing the progressive decoherence of the “meter” in a quantum measurement. *Phys. Rev. Lett.* **77**, 4887–4890 (1996).
8. Poyatos, J. F., Cirac, J. I. & Zoller, P. Quantum reservoir engineering with laser cooled trapped ions. *Phys. Rev. Lett.* **77**, 4728–4731 (1996).
9. Marzoli, I., Cirac, J. I., Blatt, R. & Zoller, P. Laser cooling of trapped three-level ions: designing two-level systems for sideband cooling. *Phys. Rev. A* **49**, 2771–2779 (1994).
10. Caldeira, A. O. & Leggett, A. J. Influence of damping on quantum interference: an exactly soluble model. *Phys. Rev. A* **31**, 1059–1066 (1985).
11. Walls, D. F. & Milburn, G. J. Effect of dissipation on quantum coherence. *Phys. Rev. A* **31**, 2403–2408 (1985).
12. Collett, M. J. Exact density-matrix calculations for simple open systems. *Phys. Rev. A* **38**, 2233–2247 (1988).
13. Wineland, D. J. *et al.* Experimental issues in coherent quantum-state manipulations of trapped atomic ions. *J. Res. NIST* **103**, 259–328 (1998).
14. Schneider, S. & Milburn, G. J. Decoherence in ion traps due to laser intensity and phase fluctuations. *Phys. Rev. A* **57**, 3748–3752 (1998).
15. Murao, M. & Knight, P. L. Decoherence in nonclassical motional states of a trapped ion. *Phys. Rev. A* **58**, 663–669 (1998).
16. Schneider, S. & Milburn, G. J. Decoherence and fidelity in ion traps with fluctuating trap parameters. *Phys. Rev. A* **59**, 3766–3774 (1999).
17. Glauber, R. J. Coherent and incoherent states of the radiation field. *Phys. Rev.* **131**, 2766–2788 (1963).
18. Raizen, M. G., Gilligan, J. M., Bergquist, J. C., Itano, W. M. & Wineland, D. J. Ionic crystals in a linear Paul trap. *Phys. Rev. A* **45**, 6493–6501 (1992).
19. Monroe, C. *et al.* Resolved-sideband Raman cooling of a bound atom to the 3D zero-point energy. *Phys. Rev. Lett.* **75**, 4011–4014 (1995).
20. Meekhof, D. M., Monroe, C., King, B. E., Itano, W. M. & Wineland, D. J. Generation of nonclassical motional states of a trapped atom. *Phys. Rev. Lett.* **76**, 1796–1799 (1996).
21. Myatt, C. J. *et al.* in *Laser Spectroscopy 13* (ed. Blatt, R., Eschner, J., Leibfried, D. & Schmidt-Kaler, F.) (World Scientific, Singapore, in the press).
22. Raimond, J. M., Brune, M. & Haroche, S. Reversible decoherence of a mesoscopic superposition of field states. *Phys. Rev. Lett.* **79**, 1964–1967 (1997).
23. Wilkinson, S. R. *et al.* Experimental evidence for non-exponential decay in quantum tunnelling. *Nature* **387**, 575–577 (1997).
24. Thompson, R. C., Hernandez-Pozos, J.-L., Höffges, J., Segal, D. M. & Vincent, J. R. in *Trapped Charged Particles and Fundamental Physics* (eds Dubin, D. H. E. & Schneider, D.) 388–392 (American Institute of Physics Conf. Proc. 457, AIP Press, Woodbury, New York, 1999).
25. Itano, W. M., Heizen, D. J., Bollinger, J. J. & Wineland, D. J. Quantum Zeno effect. *Phys. Rev. A* **41**, 2295–2300 (1990).
26. Chapman, M. S. *et al.* Photon scattering from atoms in an interferometer: coherence lost and regained. *Phys. Rev. Lett.* **75**, 3783–3787 (1995).
27. Dürr, S., Nonn, T. & Rempe, G. Fringe visibility and which-way information in an atom interferometer. *Phys. Rev. Lett.* **81**, 5705–5709 (1998).
28. Itano, W. M. *et al.* Quantum harmonic oscillator state synthesis and analysis. *Proc. SPIE* **2995**, 43–55 (1997).
29. de Matos Filho, R. L. & Vogel, W. Even and odd coherent states of the motion of a trapped ion. *Phys. Rev. Lett.* **76**, 608–611 (1996).
30. Garraway, B. M., Knight, P. L. & Plenio, M. B. Generation and preservation of coherence in dissipative quantum optical environments. *Phys. Scr.* **T76**, 152–158 (1998).

Acknowledgements

We thank the US National Security Agency, Army Research Office, and Office of Naval Research for support. We thank P. Zoller, H. Mabuchi and W. Zurek for discussions. We thank them, D. Leibfried, M. Rowe, D. Sullivan and M. Lombardi for comments on the manuscript.

Correspondence and requests for materials should be addressed to D.J.W. (e-mail: dwineland@nist.gov).

Direct protein–protein coupling enables cross-talk between dopamine D5 and γ -aminobutyric acid A receptors

Fang Liu^{*†‡}, Qi Wan[§], Zdenek B. Pristupa^{*†}, Xian-Min Yu^{*†¶}, Yu Tian Wang[§] & Hyman B. Niznik^{*†‡||}

Department of ^{*}Psychiatry, ^{||}Pharmacology, [¶]Oral Physiology and [‡]Institute of Medical Sciences, University of Toronto, Toronto, Ontario M5S-1A8, Canada

[§]Program in Brain and Behavior and Division of Pathology, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada

[†]Molecular Neurobiology Section, Center for Addiction and Mental Health, Toronto, Ontario M5T 1R8, Canada

GABA_A (γ -aminobutyric-acid A) and dopamine D1 and D5 receptors represent two structurally and functionally divergent families of neurotransmitter receptors. The former comprises a class of multi-subunit ligand-gated channels mediating fast interneuronal synaptic transmission, whereas the latter belongs to the seven-transmembrane-domain single-polypeptide receptor superfamily that exerts its biological effects, including the modulation of GABA_A receptor function, through the activation of second-messenger signalling cascades by G proteins. Here we show that GABA_A-ligand-gated channels complex selectively with D5 receptors through the direct binding of the D5 carboxy-terminal domain with the second intracellular loop of the GABA_A γ 2(short) receptor subunit. This physical association enables mutually inhibitory functional interactions between these receptor systems. The data highlight a previously unknown signal transduction mechanism whereby subtype-selective G-protein-coupled receptors dynamically regulate synaptic strength independently of classically defined second-messenger systems, and provide a heuristic framework in which to view these receptor systems in the maintenance of psychomotor disease states.

Of the numerous classes of cell-surface-receptor signalling molecules, synaptic transmission between individual neurons is mediated largely by two major structurally and functionally distinct neurotransmitter receptor families, namely ligand-gated channels and G-protein-coupled receptors (GPCRs). Although both are integral membrane proteins, ligand-gated receptors modulate synaptic neurotransmission directly through the formation and opening of an inherent ion channel, whereas GPCRs are single-polypeptide proteins containing seven hydrophobic transmembrane domains that transduce extracellular neurotransmitter signals into the cell interior by interacting with heterotrimeric G proteins^{1–3}. These in turn modulate a diverse array of cellular effectors to produce changes in cellular second-messenger systems and/or ionic conductance, and ultimately physiological responsiveness.

The GABA_A receptor is a member of the ligand-gated receptor family. After the opening of bicuculline-sensitive Cl[–] channels, GABA_A receptors mediate fast inhibitory synaptic transmission at the vast majority of inhibitory synapses in the brain, thereby taking a fundamental role in brain physiology and pathology^{4,5}. GABA_A receptors are thought to be heteropentameric structures, assembled by combining homologous subunits from five different classes⁶: α (1–6), β (1–4), γ (1–3), δ and ϵ (1 and 2). Each of the subunits contains four α -helical hydrophobic transmembrane domains. The subunit composition of the GABA_A receptors critically affects the function and pharmacology of the receptor^{7,8}. The co-expression of α and β subunits is the minimum requirement for the cell-surface expression of functional GABA_A receptors in heterologous cell lines; the full pharmacological profile of native GABA_A receptors requires the presence of α , β and γ subunits. The precise stoichiometry and subunit composition of native GABA_A receptors are still a matter of uncertainty, but the most abundant population of native neuronal GABA_A receptors is believed to be composed of α_1 , β_2 and γ_2 receptor subunits^{7–9}.

Dopamine receptors, in contrast, belong to the superfamily of GPCRs and are encoded by five distinct genes. On the basis of deduced amino-acid sequence homologies and expressed pharmacological and functional profiles these are classified further into two

subfamilies, termed D1-like and D2-like receptors¹⁰. Members of the D2-like receptor family, which include D2, D3 and D4 receptors, display a complex pattern of signal transduction, modulating diverse effector and second-messenger system pathways, as a result of their coupling to subtype-specific members of the G_i/G_o protein family. Members of the D1-like receptor family^{11,12}, referred to as D1 and D5 or D1A and D1B receptors, preferentially couple to G_s proteins, stimulating the activity of adenylyl cyclase and protein-kinase-A (PKA)-dependent pathways, whereas native D1-like receptor stimulation in both the brain and periphery is known to activate other second-messenger systems as well¹³. These receptor-mediated events might be transduced by D1 or D5 receptor interactions with G proteins other than G_s (ref. 14) or by still uncloned D1-like receptor variants. However, despite inherent differences in D1 and D5 receptor primary structures and constitutive activity profiles, clearly defined functional roles differentiating D1 receptors from D5 receptors have yet to be firmly established¹⁵.

The major intracellular loops of GABA_A receptors contain many consensus sites for protein phosphorylation, and GABA_A-receptor-mediated currents are modified by a variety of extracellular and intracellular factors after the activation of PKA-, PKC- and tyrosine-kinase-dependent pathways^{16–18}. The reported modulation of GABA_A-receptor-mediated synaptic activity by dopamine D1-like receptor stimulation is similarly thought to only occur after the activation of receptor-associated intracellular cyclic AMP protein-kinase-dependent pathways and phosphatases^{19–22}. Because there are no specific ligands that can functionally differentiate dopamine D1 from dopamine D5 receptors, the specific D1-like receptor subtype mediating the modulation of synaptic GABA_A receptor function remains unclear. In certain cortical and hippocampal neural populations, however, dopamine D5 receptors, unlike D1 receptors, are preferentially targeted to dendritic shafts and the cell soma/axon hillock area²³, where inhibitory GABAergic neurons form major postsynaptic contacts^{7,23–25}. The proposed overlap in the subcellular distribution profiles of these receptors suggests potentially specific functional interactions between D5 and GABA_A receptors. A direct test of this hypothesis has not been made.

Here we present evidence that dopamine D5 receptors, but not D1, complex with GABA_A receptors in hippocampal neurons and co-transfected cells through the direct binding of the D5 C-terminal domain with the second intracellular loop (IL2) of the GABA_A receptor γ_2 subunit. Agonist-dependent dopamine D5 and GABA_A γ_2 receptor complex formation was found to be obligatory for the functional expression and maintenance of reciprocal receptor cross-talk, as indexed by dopamine D5, but not D1, attenuated GABA_A $\alpha_1\beta_2\gamma_2$ receptor-mediated whole-cell or miniature inhibitory

postsynaptic currents, and GABA_A receptor-stimulated reductions of D5, but not D1, cAMP accumulation. These results suggest a novel method by which GPCRs and ligand-gated channels mutually regulate each other's functional status in the control of synaptic efficacy and provide a mechanistic basis for the functional differentiation of dopamine D5 and D1 receptors.

D5 and GABA_A γ_2 subunit direct interactions

D1 and D5 receptors display particular sequence divergence within

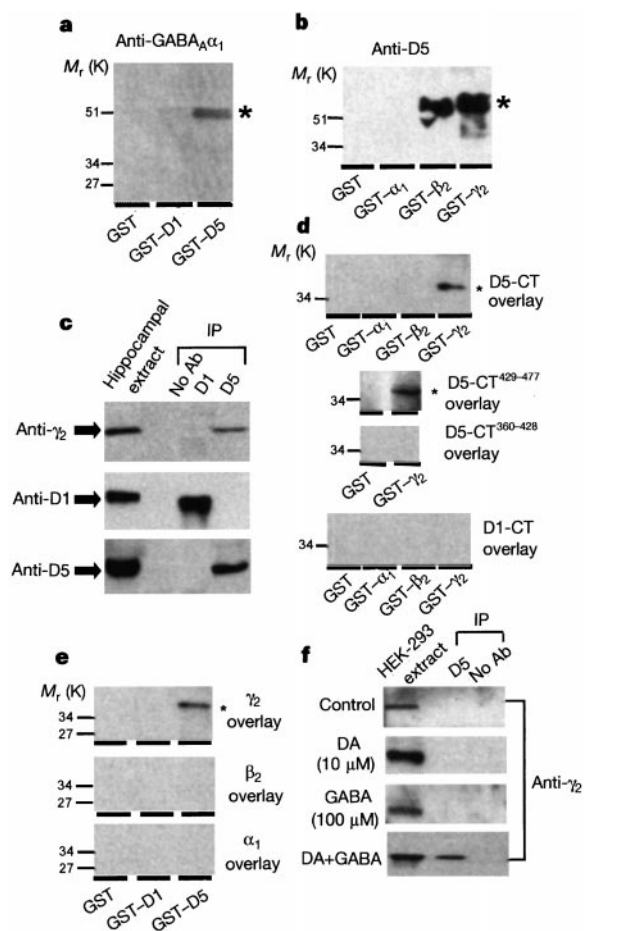


Figure 1 Association of D5 and GABA_A receptors in rat brain and *in vitro*. **a**, **b**, Western blots of hippocampal **(a)** GABA_A receptor α_1 subunit after affinity precipitation by D5-CT³⁶⁰⁻⁴⁷⁷ but not by D1-CT³³²⁻⁴⁴⁶ GST fusion proteins and **(b)** D5 receptors after incubation with GST fusion protein of GABA_A receptor second intracellular loop α_1 ^{N334-R420}, β_2 ^{N327-R451} and γ_2 ^{H322-S442} (short form) subunits or GST alone. **c**, Co-immunoprecipitation of hippocampal GABA_A receptors by D5 but not D1 receptor antibody (Ab). **d**, Filter overlay assays depicting the direct binding of the D5-CT domain to the second intracellular loop of the γ_2 subunit. SDS-PAGE-separated GST fusion proteins encoding the second intracellular loop of α_1 , β_2 and γ_2 subunits were probed with *in vitro* translated ³⁵S-labelled D5-CT (top) or ³⁵S-labelled D1-CT peptides (bottom). Amino-acid sequence D5-CT⁴²⁹⁻⁴⁷⁷ was critical for binding to γ_2 IL-GST (middle). **e**, Direct binding of the γ_2 intracellular loop to D5 receptor C-terminal tail sequences. GST fusion proteins of D1 or D5-CT probed with ³⁵S-labelled γ_2 (top), β_2 (middle) or α_1 (bottom) second intracellular loops. GST was used as a control. The positions of molecular size markers are indicated at the left. **f**, Co-immunoprecipitation (IP) of GABA_A receptors by D5 receptor antibodies in co-expressing HEK-293 cells is dependent on agonist co-activation of both receptors. Data are representative of two to six independent experiments.

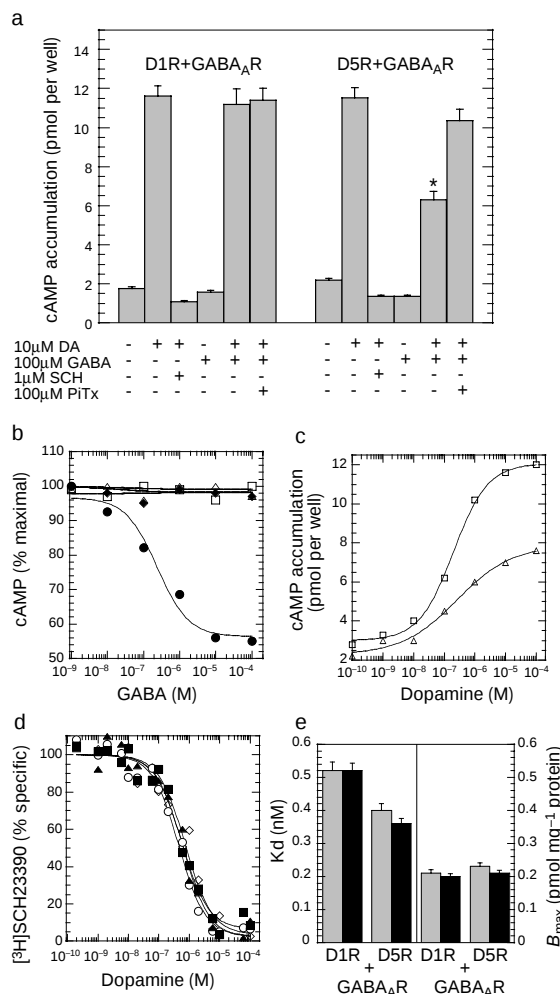


Figure 2 Selective modulation of dopamine-mediated D5 receptor-stimulated cAMP accumulation by GABA_A $\alpha_1\beta_2\gamma_2$ receptors in co-expressed HEK-293 cells. **a**, GABA pretreatment significantly reduces maximal dopamine (DA)-stimulated cAMP accumulation by D5, but not D1, receptors; this effect is blocked by the GABA_A antagonist picrotoxin (Ptx). Suffix R, receptor; SCH, SCH-23390. **b**, Concentration-dependent reduction of D5 maximal dopamine (10 μ M)-stimulated cAMP accumulation by GABA. Squares, D1 receptor; triangles, D5 receptor; diamonds, D1 and GABA_A receptors; circles, D5 and GABA_A receptors. **c**, GABA_A receptor stimulation reduces the efficacy (0.6 versus 1.0) but not the affinity of dopamine D5 receptor-stimulated cAMP accumulation. Estimated EC₅₀ and V_{max} values are listed in the text. Squares, control; triangles, treatment. **d**, The affinity of dopamine for D5 receptors is unchanged after GABA treatment, as indexed by [³H]SCH-23390 (350 pM final concentration) binding. K_d values are listed in the text. Diamond, D5 receptor; circles, D5 receptor with 0.1 mM GABA; squares, D5 and GABA_A receptors; triangles, D5 and GABA_A receptors with 0.1 mM GABA. **e**, GABA (100 μ M) treatment (hatched columns) did not alter the estimated K_d or B_{max} of [³H]SCH-23390 binding. Grey columns, control. Data are representative as means $\pm$ s.e.m. of three to eight independent experiments, each conducted in duplicate. Suffix R, receptor. Asterisk, $P < 0.05$ (Student's two-tailed *t*-test).

the C-terminal tail, a region that might confer subtype-selective functional attributes and coupling to distinct effectors independent of those classically associated with G-protein activation²⁶. To investigate potential interactions between D5 and GABA_A receptors, we incubated rat hippocampal brain extracts with glutathione S-transferase (GST) fusion proteins encoding the D5 C-terminal tail (D5-CT) or the D1 C-terminus (as control). As illustrated in Fig. 1a, a band of relative molecular mass 51,000 (M_r 51K) that immunoreacted with the GABA_A receptor α_1 subunit was precipitated with GST–D5-CT^{A360–H477}-incubated hippocampal membranes but not with GST–D1-CT^{A332–T446} or GST alone, indicating that the native D5 receptor might complex with the GABA_A receptor through its C-terminal tail *in situ*. To identify the GABA_A subunit responsible for this interaction, we constructed GST fusion proteins encoding the major intracellular domains (IL2) of α_1 ^{N334–R420}, β_2 ^{N327–R451} and γ_2 ^{H522–S442} GABA_A receptor subunits, the most predominant molecular form of GABA_A receptor expressed in the hippocampus^{7,8,27}. β_2 or γ_2 GST–IL fusion proteins, but not GST– α -IL, precipitated solubilized hippocampal D5 (M_r 60K), but not D1, receptors

(Fig. 1b). The physical association of D5 and GABA_A receptors was confirmed by immunoprecipitation experiments in which D5, but not D1, receptor antibodies precipitated hippocampal GABA_A receptors (Fig. 1c), indicating relevant complex formation between these two holoreceptor moieties *in vivo*.

To assess whether D5–GABA_A receptor complex formation was mediated by the direct binding of D5-CT to GABA_A receptor β_2 and/or γ_2 subunits or, indirectly, by interaction with receptor accessory proteins^{28,29}, we probed, in blot overlay experiments, GST fusion proteins encoding IL2 of α_1 , β_2 and γ_2 GABA_A receptor

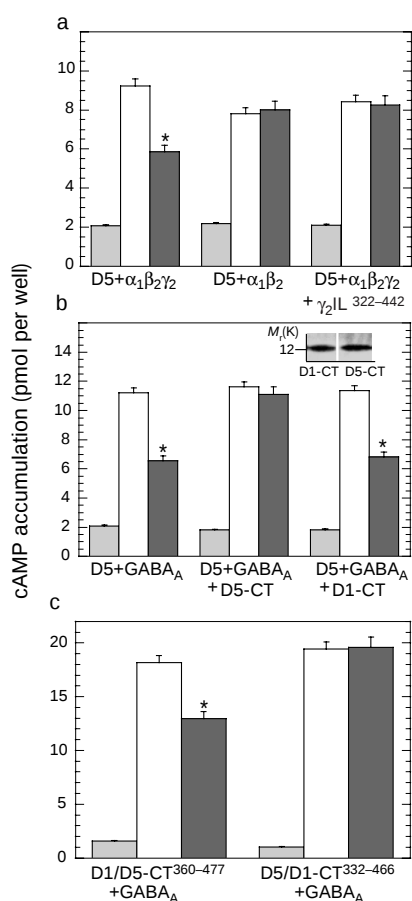


Figure 3 GABA_A receptor modulation of D5 cAMP production is dependent on γ_2 subunit and D5-CT sequences in co-expressed cells. Columns in all panels as follows: stippled, basal; white, dopamine; hatched, GABA and dopamine. **a**, GABA did not reduce D5 receptor-mediated cAMP accumulation in GABA_A $\alpha_1\beta_2$ -expressing cells, indicating the requirement for the γ_2 subunit. Co-expression of a minigene encoding γ_2 ^{322–442} subunit amino acids inhibited GABA_A $\alpha_1\beta_2\gamma_2$ receptor modulation of D5 cAMP accumulation, as do minigenes encoding D5-CT^{360–477} but not D1-CT^{332–446} sequences (**b**). Inset, western blot of expressed D1-CT and D5-CT peptides. **c**, Chimaeric D5 receptors in which the C-terminal tail domain was swapped with sequences encoded by dopamine D1-CT (D5/D1-CT^{332–446}) are unresponsive to GABA_A $\alpha_1\beta_2\gamma_2$ receptor stimulation, whereas D1/D5-CT^{360–477} receptor mutants fully reconstitute GABA_A-receptor-mediated decreases in cAMP activity. Data are means $\pm$ s.e.m. of three to six independent experiments, each conducted in duplicate. Asterisk, $P < 0.05$.

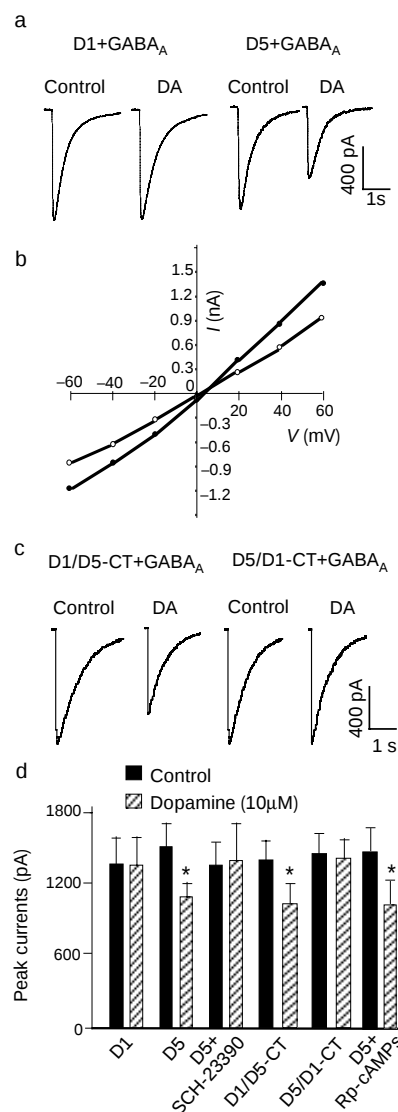


Figure 4 D5 receptor regulation of GABA_A-receptor-mediated whole-cell currents in co-transfected HEK-293 cells. **a**, D5 receptor, but not D1 receptor, stimulation by dopamine (10 μ M) reduces GABA_A receptor-mediated whole-cell currents by $\sim 30\%$ ($n = 6$, $P < 0.01$), an event blocked by the dopamine (DA) D1-like receptor antagonist SCH-23390 ($n = 6$). **b**, I/V relationship of GABA_A receptor currents before (filled circles) and after D5 receptor stimulation (dopamine, 10 μ M; open circles) at holding potentials of -60 to $+60$ mV. Dopamine does not change the reversal potential of the currents but decreases the slope of the current–voltage curve. **c**, Modulation of GABA_A-receptor-mediated currents are dependent on dopamine D5-CT sequences. Chimaeric D5/D1-CT^{332–446} receptors failed to reduce the GABA_A receptor-mediated currents, whereas D1/D5-CT^{360–477} receptor mutants reduced currents. **d**, Bar graph depicting the mean peak GABA_A current values under the conditions described and after pretreatment with the PKA inhibitor Rp-cAMPS. Values are means $\pm$ s.e.m. from six independent experiments ($P < 0.01$). Black columns, control; hatched columns, dopamine (10 μ M).

subunits with *in vitro* translated [35 S]methionine-labelled D5-CT and D1-CT peptides (35 S-D5-CT and 35 S-D1-CT), respectively. As depicted in Fig. 1d (top), the 35 S-D5-CT probe binds, and interacts specifically with, IL2 of the γ_2 GABA $_A$ receptor subunit, but not with control GST or GST fusion proteins expressing either α_1 or β_2 IL2 GABA $_A$ receptor subunits, and seems dependent on the sequence encoded by the last 49 amino acids (429–477) of the D5-CT (Fig. 1d, middle). The 35 S-D1-CT probe does not recognize GST fusion proteins with GABA $_A$ α_1 -, β_2 - or γ_2 -subunit intracellular loops (Fig. 1d, bottom). Conversely, as shown in Fig. 1e (top), *in vitro* translated [35 S]methionine-labelled IL2 peptides of γ_2 (35 S- γ_2 IL), but not 35 S- β_2 IL (middle) or 35 S- α_1 IL (bottom) subunits, bound GST–D5-CT fusion proteins but failed to bind to GST alone or GST–D1-CT. These results suggest that hippocampal D5 receptors complex with GABA $_A$ receptors, presumably by direct binding of the D5-CT tail to the γ_2 intracellular loop. The fact that GST fusion proteins expressing the GABA $_A$ receptor β_2 subunit intracellular loop can affinity-precipitate native D5 receptors (Fig. 1b), but do not recognize D5-CT sequences in blot overlay assays, would suggest either that the β_2 subunit binds directly to D5 receptor sequences outside the C-terminal tail domain or, alternatively, that an indirect protein–protein complex between β_2 subunits and D5 receptors exists.

Agonist dependence of D5– γ_2 complex formation

We next ascertained whether D5 and GABA $_A$ receptor protein–protein binding is dynamically regulated in HEK-293 cells. Under conditions where D1 or D5 receptor expression levels were similar (see Fig. 2e) and transiently co-expressed with GABA $_A$ α_1 , β_2 and γ_2 receptor subunits, the formation of complexes between dopamine D5 and the GABA $_A$ receptor γ_2 subunit was found to be dependent on receptor co-activation by their respective agonists. As illustrated in Fig. 1f, immunoprecipitation of GABA $_A$ –D5 receptor complexes by D5 receptor antibodies is evident only in cells that were incubated with both dopamine and GABA but not after agonist stimulation of either GABA $_A$ or dopamine D5 receptors alone. Pretreatment of cells with either the D1/D5 receptor antagonist SCH-23390 (10 μ M) or the GABA $_A$ receptor blocker picrotoxin (100 μ M) alone before receptor co-activation by agonists prevented the formation of a GABA $_A$ –D5-receptor complex (data not shown).

Reciprocal receptor cross-talk through D5CT– γ_2 complexes

The regulation of GABA $_A$ –D5 complex formation by agonist co-occupied receptors suggests the existence of functionally relevant interactions between these receptors. In co-expressing cells, dopamine (10 μ M) maximally stimulated the production of cAMP by either D5 or D1 receptors to virtually identical extents, an effect that was blocked by the D1 and D5 receptor antagonist SCH-23390. When cells were pretreated briefly with 100 μ M GABA, the maximal stimulation of D5-mediated adenylyl cyclase activity by dopamine was decreased by $\sim$ 45%, whereas there were no changes in D1-receptor-mediated cAMP levels (Fig. 2a). GABA-mediated decreases in D5-receptor-stimulated cAMP accumulation were blocked by the GABA $_A$ receptor antagonist picrotoxin (100 μ M), indicating modulation by activation of GABA $_A$ receptors. GABA inhibited D5-receptor-stimulated cAMP accumulation in a concentration-dependent manner (estimated concentration for half-maximal response (EC_{50}) 292 ± 40 nM) but exerted no effect on dopamine-mediated D1-receptor-stimulated cAMP accumulation in cells in which they were expressed together (Fig. 2b). GABA did not modulate either dopamine-mediated D5- or D1-receptor-stimulated cAMP accumulation in cells expressing these receptors alone (Fig. 2b). The decrease in maximal dopamine-mediated D5-receptor-stimulated cAMP production cannot be attributed to either a GABA-induced competitive attenuation of agonist D5 receptor affinity or decreased receptor expression levels, as (1) the estimated EC_{50} for dopamine-mediated D5-receptor-stimulated

cAMP production in GABA-pretreated cells (302 ± 43 nM) was virtually identical to that observed in control cells (250 ± 38 nM) despite a $\sim$ 40% decrease in apparent maximal accumulation of cAMP (Fig. 2c; $n = 3$); (2) the estimated dissociation constant (K_i) for dopamine at D5 receptors after GABA pretreatment (308 ± 46 nM) was not significantly different from that of control (349 ± 51 nM) or untreated (289 ± 31 nM) GABA $_A$ -receptor-expressing cells (Fig. 2d; $n = 4$); and (3) GABA treatment did not significantly modify either the estimated dissociation constant (K_d) (Fig. 2e, left) or maximal binding (B_{max}) (Fig. 2e; right) of [3 H]SCH-23390 to either D1 or D5 receptors.

The selective decrease in dopamine-mediated D5-receptor-stimulated, but not D1-receptor-stimulated, cAMP accumulation by GABA $_A$ receptor activation suggested that D5–GABA $_A$ γ_2 receptor complexes might underlie this event. We obtained evidence in support of this contention by determining the GABA $_A$ receptor subunit requirement for this interaction. As illustrated in Fig. 3a, the modulation of D5-receptor-mediated cAMP accumulation by GABA $_A$ receptors was completely abolished when only GABA $_A$ α_1 and β_2 subunits were co-expressed. Because the γ_2 subunit is not required for the formation of functional GABA $_A$ receptor channels in the expression system employed here^{30,31}, the failure of GABA $_A$ receptors lacking this subunit to modulate D5-receptor-mediated cAMP accumulation seems to be independent of channel-mediated activation of intracellular kinases or phosphatases and is consistent with the proposed mechanism involving the selective coupling of D5 receptors to γ_2 subunits. Corroborating evidence was obtained with the use of minigenes encoding sequences of either IL2 of the γ_2 GABA $_A$ receptor subunit or D5-CT. We reasoned that in cells co-expressing D5 and GABA $_A$ α_1 , β_2 and γ_2 receptor subunits, the additional expression of either γ_2 intracellular loop or D5-C-terminal sequence motifs would act as competitive inhibitors³² of putative D5-CT GABA $_A$ - γ_2 subunit protein recognition domains. As depicted in Fig. 3a, minigenes encoding peptides of IL2 of GABA $_A$ receptor γ_2 subunit completely inhibited the modulation of D5 cAMP accumulation by GABA $_A$ receptor ($n = 4$), whereas those encoding IL2 of GABA $_A$ α_1 or β_2 receptor fragments were without effect (data not shown). Conversely, minigenes encoding human D5-CT inhibited the agonist-promoted ability of GABA $_A$ receptors to decrease D5-receptor-stimulated cAMP accumulation, whereas GABA $_A$ -receptor-mediated decreases in D5-receptor-stimulated cAMP accumulation ($P < 0.05$, $n = 3$) were unaffected in cells expressing the human dopamine D1-CT peptide (Fig. 3b). In control cells, neither D5 nor D1-CT peptides modulated dopamine-mediated D1- or D5-receptor-stimulated cAMP accumulation (data not shown).

To confirm that the observed modulation of D5-receptor-mediated responses by GABA $_A$ receptors is a sole product of the agonist-promoted direct protein–protein binding between D5-CT and intracellular loop γ_2 amino-acid motifs, we constructed chimaeric human D1 and D5 receptors in which the CT tails were

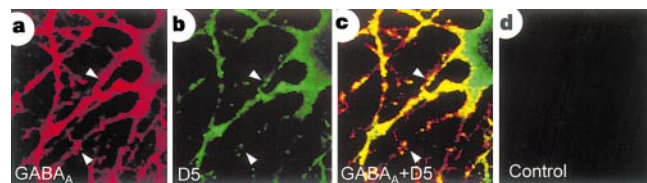


Figure 5 Co-localization of GABA $_A$ and dopamine D5 receptors in cultured hippocampal neurons. **a, b**, Confocal optical sections of the distribution of GABA $_A$ receptors (**a**) and D5 receptors (**b**) in cultured hippocampal neurons was obtained by double immunostaining with bd-17 and D5 antibodies. Dopamine D5 receptors were revealed by FITC-conjugated secondary antibodies (green) and GABA $_A$ receptors by Cy3 (red). **c**, Superimposition of confocal images in **a** and **b**. Arrowheads highlight areas of punctate GABA $_A$ and D5 receptor clustering. No staining was observed when primary antibodies were eliminated (**d**) or when antibodies were preabsorbed with antigen (data not shown).

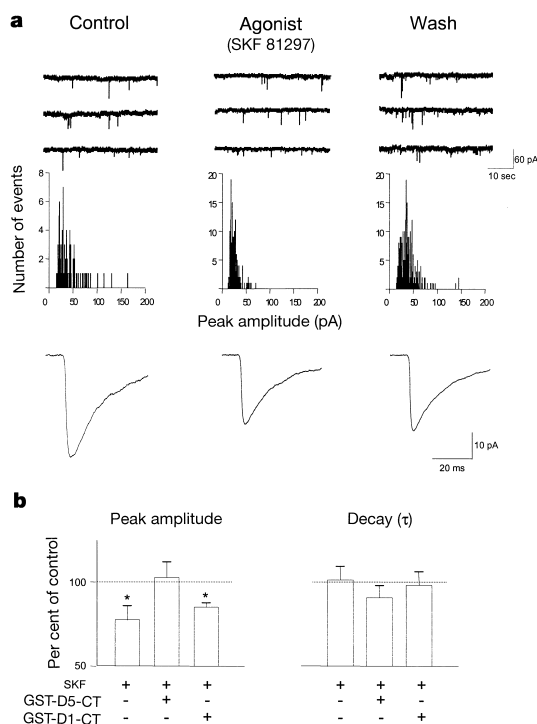


Figure 6 Dopamine D5 receptors modify GABA_A-receptor-mediated synaptic responses in hippocampal neurons. **a**, Example of GABA_A-receptor-mediated mIPSCs recorded before (control) or during the application of the D1/D5 receptor agonist SKF-81297 (10 μ M) and after agonist wash-out. The top panel shows representative recording traces, the middle panel amplitude distribution histograms of the mIPSCs recorded and

the bottom panel traces of averaged mIPSCs. **b**, In the presence of both PKC and PKA inhibitors, stimulation by D1/D5 receptor agonist significantly decreases the amplitude of GABA_A-receptor-mediated mIPSCs ($n = 4$); this effect can be prevented by the intracellular infusion of GST-D5-CT fusion protein into recorded neurons ($n = 5$), but not by GST-D1-CT fusion protein ($n = 7$, **b**). Asterisk, $P < 0.05$ (Student's paired t -test).

swapped. These receptors have been characterized and determined to be fully functional in expressed cells³³ (L. Demchyshyn, F. McConkey and H.B.N., manuscript in preparation). As illustrated in Fig. 3c, D1 receptor mutants, in which the C-terminal tail was replaced by D5-CT-encoded sequences, fully reconstituted the ability of GABA_A receptor stimulation to reduce D1/D5-CT-mediated cAMP accumulation by 30% ($P \leq 0.05$, $n = 4$), whereas D5 receptors in which the C-terminal tail was replaced with D1-CT sequences failed to establish functional interactions with GABA_A receptors. These data strongly suggest that agonist-stimulated D5-CT- γ_2 complex formation is sufficient and required for the expression of the modulation of D5 functional responses by GABA_A receptors.

We next assessed whether the physical interaction between these two receptors enables the modulation of GABA_A receptor function after activation of D5 receptors, but not D1 receptors, by examining the effect of dopamine on GABA_A-receptor-mediated whole-cell currents in HEK cells. Application of dopamine in a bath produced a decrease in GABA_A receptor-mediated currents by $\sim 30\%$ in cells co-transfected with D5 receptors, but failed to modify currents in cells expressing D1 receptors (Fig. 4a) despite comparable levels of expressed D1 and D5 receptor densities (see also Fig. 2). The D5-induced inhibition of GABA_A currents is a result of a decreased slope of the current-voltage curve with no change in the reversal potential (Fig. 4b). This inhibition was concentration-dependent with an estimated EC_{50} of 250 ± 34 nM, virtually identical to values for dopamine-mediated D5-receptor-stimulated cAMP accumulation in these cells (Fig. 2d) and was completely blocked by the D1-like-receptor antagonist SCH-23390 (Fig. 4d). Control cells expressing only GABA_A receptors were completely unresponsive to dopamine. Thus, dopamine D5, but not D1, receptor activation selectively downregulates the function of GABA_A receptors.

The failure of dopamine-mediated D1 receptor stimulation to modify GABA_A receptor currents strongly indicates that the D5-

receptor-mediated attenuation of GABA_A receptor function occurs through a novel cAMP-independent mechanism. This notion is supported by the inability of the potent PKA inhibitor Rp-cAMPs (10 μ M) to modify D5-receptor-mediated decreases in GABA_A receptor-mediated whole-cell currents (Fig. 4d) (1449 ± 146 pA (control); 1004 ± 126 pA (treated); $n = 6$, $P < 0.01$) and is consistent with observations that β_2 -subunit-containing GABA_A receptors are insensitive to post-translational modification by intracellular cAMP³⁴. Moreover, we show that the selective modulation of GABA_A-receptor-mediated currents by D5 receptor stimulation were entirely dependent on sequence motifs of the C-terminal tail (Fig. 4c, d). Thus, D1 receptor substitution mutants expressing D5-CT sequences fully reconstitute receptor cross-talk with GABA_A receptors, inhibiting peak currents by 30%, whereas the activation of D5 receptors in which the C-terminal tail was substituted with D1-CT sequences failed to modulate GABA_A-receptor-mediated currents. Together, these data suggest that agonist-promoted direct binding of D5-CT sequences to GABA_A receptor γ_2 subunit are obligatory for the expression of functional reciprocal D5 and GABA_A receptor cross-talk.

D5 and GABA_A receptor interactions in neurons

The proposed direct protein-protein interaction between D5 and GABA_A receptors described above requires that these receptor moieties be co-expressed within the same neuron. Although reports of the subcellular distribution profiles of D5 receptors²³ seem to parallel and be consistent with the reported cellular and subcellular distributions of GABA_A receptors in certain cortical and hippocampal cell populations^{24,25,27}, direct experimental evidence for D5 and GABA_A receptor co-localization is lacking. Figure 5 depicts the immunofluorescence and subcellular co-localization patterns of D5 and GABA_A receptors within cultured hippocampal neurons. The strong overlap in the clustering of D5 and GABA_A receptor moieties in these hippocampal neurons suggests that D5 receptors,

by virtue of their close spatial proximity to GABA_A receptors, can serve a distinct function by preferentially modulating inhibitory synaptic inputs.

To determine whether the proposed complex formation between D5 and GABA_A receptors is physiologically relevant in regulating synaptic strength *in situ*, we examined the ability of D1 and D5 receptors to modulate GABA_A-receptor-mediated miniature inhibitory postsynaptic currents (mIPSCs) in cultured hippocampal neurons under conditions where both PKA- and PKC-dependent pathways were blocked by application of Rp-cAMPs (10 μM) and staurosporine (1 μM) into recorded neurons. Whereas application of the D1 and D5 receptor agonist SKF-81297 (10 μM) in a bath significantly increased bicuculline (10 μM)-sensitive (data not shown) GABA_A mIPSC frequency (Fig. 6a), indicating enhanced presynaptic GABA release, the peak amplitude of the mIPSCs was decreased to 77 ± 8% of control (32.4 ± 5.3 versus 24.1 ± 3.1 pA (SKF-81297-treated); *n* = 4). No significant changes in either the rise (10–90% of the peak, 2.5 ± 0.6 versus 2.1 ± 0.5 ms; *n* = 4) or the decay time constant (*τ*, 34.2 ± 5.1 versus 34.6 ± 4.6 ms; *n* = 4) were observed. Application of GST-encoded D5-CT fusion protein into neurons through the recording pipette prevented the decrease in mIPSCs induced by application of SKF-81297 (*n* = 5) (Fig. 5b), whereas GST–D1-CT fusion peptides had no such effect (*n* = 7) (Fig. 6b). Neither peptide modulated the increased frequency of GABA_A mIPSCs induced by SKF-81297 (data not shown). These data support the contention that complex formation between GABA_A and D5 receptors can modulate synaptic currents *in vivo*.

Discussion

A number of GPCRs have been documented to form either homodimeric³⁵ or heterodimeric protein–protein interactions^{36,37} with subtype-specific variants of the same receptor subclass to permit the full expression of receptor-specific pharmacological and functional attributes. We provide data describing a mechanism by which two structurally and functionally distinct neurotransmitter receptor classes, as represented by multimeric ligand-gated ion channels and single-polypeptide G-protein-linked proteins, are coupled. Agonist-dependent protein–protein complex formation between the C-terminal-tail domain of D5 receptors and IL2 of GABA_A receptors γ₂ subunits permits the rapid, reciprocal modulation of GABA_A α₁β₂γ₂ and dopamine D5-receptor-mediated events independently of their respective classically defined signal transduction cascades. Given the extensive co-localization of GABA_A and D5 receptors, we propose that after receptor co-activation by endogenous neurotransmitters, dopamine D5 receptors might preferentially be associated with and modulate inhibitory synaptic inputs in some hippocampal or cortical neuronal populations by means of direct protein–protein interactions with γ₂ subunits. This contention is particularly intriguing because aberrations in cortical or hippocampal GABAergic γ₂ (refs 38–40) and dopaminergic D1-like receptor activity⁴¹, possibility as a result of cortical hypofunctionality, has been postulated to have a key role in the pathophysiology of both the positive and negative symptomatology of schizophrenia^{42,43}. The data presented here might provide a heuristic framework in which to view these co-directional deficits and indicate a potential therapeutic avenue for psychomotor and neuropsychiatric disease states known to be defective in both GABAergic and dopaminergic tone.

The exact molecular basis for the observed co-directional decrease in GABA_A and D5 receptor function is unknown. Unlike that of functional D1-like-receptor-mediated events⁴⁴, the magnitude of GABA_A-receptor-mediated inhibitory postsynaptic responses seems to be dependent on the number of receptors available on the cell surface^{45,46}. It is known that, in addition to binding the C-terminal of D5 receptors, the γ₂ subunit is also required for the postsynaptic clustering of GABA_A receptors, possibly through interaction with gephyrin^{28,47} and other receptor-

associated proteins²⁹. Thus, the agonist-dependent formation of γ₂–D5-CT complexes might influence the localization and/or cell-surface trafficking of neuronal GABA_A receptors.

Despite the implied and observed functional redundancy of multiple dopamine D1-like receptors with respect to expressed pharmacological profiles and agonist-mediated second-messenger-system activation¹³, highly variable amino-acid sequence motifs with the C-terminal tail can confer exquisite subtype-selective receptor functions. It remains to be investigated whether direct protein–protein interaction represents an alternative means of signal transduction restricted to GABA_A and D5 receptor subtypes, or whether it is indicative of a much more widespread phenomenon between numerous other classes of co-localized neurotransmitter receptor systems. □

Methods

Chimaeric receptors, GST fusion proteins and minigenes

Mutant D1/D5-CT^{360–477} and D5/D1-CT^{332–446} receptors were generated by a two-step polymerase chain reaction (PCR) using gene splicing by overlap extension with anchored oligonucleotide primers (HSC) directed to the conserved amino-acid sequence motif YAFNADF (one-letter codes), after transmembrane domain 7, using methods and procedures described previously³⁵. Purified full-length DNA (Glassmax spin columns, Gibco/BRC) was subcloned into pcDNA3. Dopamine D1, D5-CT or IL2 of α₁, β₂ or γ₂ subunit cDNA-encoding fragments were amplified by PCR. All 5' and 3' oligonucleotides incorporated *Bam*HI and *Eco*RI sites, respectively, to facilitate subcloning into pcDNA3 or pGEX2T (Pharmacia) and incorporated initiation methionine residues and stop codons where appropriate. GST fusion proteins were prepared from bacterial lysates as described by the manufacturer. To confirm appropriate splice fusion and the absence of spurious PCR-generated errors, all constructs were resequenced as described^{12,33}.

Affinity purification ('pull-down'), co-immunoprecipitation and western blotting

Rat hippocampus (5 g) or transfected human embryonic kidney (HEK)-293 cells (~2 × 10⁷) were homogenized in 50 mM Tris-HCl (pH 7.6), 150 mM NaCl, 1% IgepalCa630, 0.5% sodium deoxycholate, 2 mM EDTA, 1 mM sodium orthovanadate, 1 mM PMSF and proteinase inhibitor cocktail (Sigma, 5 μl per 100 mg tissue), centrifuged (10,000 g at 4 °C for 20 min) and membranes were solubilized with 1% Triton X-100. Hippocampal Triton extracts (~200 μg protein) were incubated with glutathione–Sepharose beads (Pharmacia) bound to indicated GST fusion proteins (~50–100 μg) at room temperature for 1 h. Beads were washed three times with 400 μl PBS containing 0.1% Triton with bound proteins, and eluted twice with 20 μl glutathione elution buffer. Eluates were incubated in sample buffer (final concentration 5% SDS) and subjected to SDS–PAGE (10% gel). Transferred GST fusion proteins were revealed by Coomassie blue staining and western blot analysis as described below. For co-immunoprecipitation experiments, hippocampal or HEK cell lysates (~500 μg protein) were incubated in the presence or absence of primary D1/D5 antibodies (2 μg) for 1 h at 4 °C, followed by the addition of 20 μl of Protein G–Sepharose (Sigma) for 12 h. Pellets were washed four times in buffer, boiled for 90 s in SDS sample buffer and subjected to SDS–PAGE. Blots were blocked with 5% non-fat dried milk dissolved in TBST buffer (10 mM Tris-HCl, 150 mM NaCl and 0.1% Tween) for 1 h at room temperature, washed three times and incubated with primary antibodies against GABA_A α₁ subunit (diluted 1:1,000, Pharmingen), D1 (1:500, RBI) or D5 (1:500, Santa Cruz) receptors for 1 h at room temperature. Rat γ₂ subunit (short form) rabbit polyclonal antibody was raised against GST fusion protein encoding the entire intracellular loop between the third and fourth transmembrane domains. This antibody, which can be competitively blocked with antigen peptide, was used at 1:750 dilution. Proteins were revealed with peroxidase-conjugated secondary antibodies (Sigma) and enhanced chemiluminescence (Amersham).

Blot overlay assays

GST fusion proteins of GABA_A receptor α₁, β₂ or γ₂ subunits or D1 or D5-CT or GST alone (50 μg each) were incubated, after transfer to polyvinylidene difluoride in 2% milk-TBST buffer⁴⁸ and [³⁵S]methionine-labelled D1 or D5 C-terminus or GABA_A IL2 α₁, β₂ or γ₂ peptide probes for 60 min at 22 °C, using TNT T7 Quick Coupled Transcription/Translation kit (Promega) incorporating 2 μl [³⁵S]methionine (1,000 Ci mmol⁻¹; 10 μCi ml⁻¹) as described by the manufacturer. Blots were washed five times with overlay buffer in PBS and subjected to autoradiography using BioMax (Kodak) film. Assays were performed three times with similar results.

Cell transfection, cAMP and ligand binding assays

Rat GABA_A receptor subunit cDNAs, pRc/CMV-α₁, pcDNA1-β₂ and pcDNA3-γ₂ (short form), were gifts from C. Kaufman and D. Gunersen (Laboratory of Neuroscience, NIH). Dopamine D1¹¹ and D5¹² clones and the cDNA encoding GFP was inserted into pcDNA3. HEK-293 cells were transfected with 10 μg each of α₁β₂γ₂ receptor subunits, D1 or D5 receptor plasmid, or, where indicated, other constructs by the calcium phosphate method. At 2–4 d after transfection, cells were assayed for cAMP accumulation by radioimmunoassay (Amersham) and for [³H]SCH-23390 (0.03–3.0 nM; 85 Ci mmol⁻¹; NEN) ligand binding activity with materials and methods as previously described^{11,12,33}.

Cell culture, electrophysiology and confocal imaging

Primary cultures from hippocampus were prepared from fetal Wistar rats (embryonic day 17–19) as described^{46,49} on Cell⁺ (Sarstedt) culture dishes. Whole-cell recordings in cultured hippocampal neurons and HEK-293 cells were made essentially as described^{31,46,49}. For mIPSC recordings the extracellular solution was further supplemented with MgCl₂ (2 mM), tetrodotoxin (1 μ M), DNQX (6 cyano-7-nitroquinoxaline-2,3-dione) (10 μ M) and APV (($\pm$)-2-amino-5-phosphonopentanoic acid) (50 μ M) to block excitatory synaptic transmission; both PKA and PKC inhibitors (Rp-cAMPs (adenosine 3'5'-cyclic monophosphothioate triethylamine) 10 μ M, staurosporine 10 μ M) and GST-encoded D5-CT or D1-CT (300 μ M) fusion peptides were included in the intracellular recording solution where indicated. GABA_A-receptor-mediated currents were recorded in a standard whole-cell voltage-clamp configuration, at a holding potential of –60 mV unless otherwise indicated, using an Axopatch 1D or 200B amplifier (Axon Instruments). Whole-cell currents in HEK-293 cells were evoked by pressure-ejection of GABA (100 μ M) from a micropipette located 20–50 μ m from the cell at 1 min intervals with dopamine or other agents added to the bath once stable recordings were obtained. Current recordings were sampled with an IBM-PC-compatible computer using pClamp software (pClamp6, Axon Instruments). mIPSC recordings in hippocampal neurons started 10 min after the formation of the whole-cell recording configuration; mIPSC data were analysed as described⁴⁹.

Immunofluorescent confocal images of the co-localization of dopamine D5 and GABA_A receptors in cultured hippocampal neurons were performed with methods described previously⁴⁶.

Received 19 July; accepted 26 October 1999.

- Dohlman, H. G., Thörner, J., Caron, M. G. & Lefkowitz, R. J. Model systems for the study of seven transmembrane segment receptors. *Annu. Rev. Biochem.* **60**, 653–688 (1991).
- Unwin, N. Neurotransmitter action: opening of ligand gated ion channels. *Cell* **72** (suppl.), 653–688 (1991).
- Neer, E. J. Heterotrimeric G proteins: organizers of transmembrane signals. *Cell* **80**, 249–257 (1995).
- McDonald, R. L. & Olsen, R. W. GABA_A receptor channels. *Annu. Rev. Neurosci.* **17**, 569–602 (1994).
- Mody, I., DeKoninck, Y., Otis, T. S. & Soltesz, I. Bridging the cleft at GABA synapses in the brain. *Trends Neurosci.* **17**, 517–525 (1994).
- Barnard, E. et al. Subtypes of GABA_A receptors: classification on the basis of subunit structure and receptor function. *Pharmacol. Rev.* **50**, 291–313 (1998).
- Costa, E. From GABA receptor diversity emerges a unified vision of GABAergic inhibition. *Annu. Rev. Pharmacol. Toxicol.* **38**, 321–350 (1998).
- McKernan, R. M. & Whiting, P. J. Which GABA_A receptor subtypes really occur in the brain? *Trends Neurosci.* **19**, 139–143 (1996).
- Tretter, V., Ehya, N., Fuchs, K. & Sieghart, W. Stoichiometry and assembly of a recombinant GABA_A receptor subtype. *J. Neurosci.* **17**, 2728–2737 (1997).
- Missale, C., Nash, S. R., Robinson, S. W., Jaber, M. & Caron, M. G. Dopamine receptors: from structure to function. *Physiol. Rev.* **78**, 189–225 (1998).
- Sunahara, R. K. et al. Human dopamine D1 receptor encoded by an intronless gene on chromosome 5. *Nature* **347**, 80–83 (1990).
- Sunahara, R. K. et al. Cloning of a gene for a human dopamine D5 receptor with higher affinity for dopamine than D1. *Nature* **350**, 614–619 (1991).
- Niznik, H. B., Sugamori, K. S., Clifford, J. J. & Waddington, J. L. in *Handbook of Experimental Pharmacology—Dopamine in the CNS*. (ed. Di Chiara, G.) (Springer, Berlin, in the press).
- Sidhu, A. Coupling of D1 and D5 dopamine receptors to multiple G proteins. *Mol. Neurobiol.* **16**, 125–134 (1998).
- Sibley, D. R. New insights into dopaminergic function using anti-sense and genetically altered animals. *Annu. Rev. Pharmacol. Toxicol.* **39**, 313–341 (1999).
- Smart, T. G. Regulation of excitatory and inhibitory neurotransmitter-gated ion channels by protein phosphorylation. *Curr. Opin. Neurobiol.* **7**, 358–367 (1997).
- Poisbeau, P., Cheney, M. C., Browning, M. D. & Mody, I. Modulation of synaptic GABA_A receptor function by PKA and PKC in adult hippocampal neurons. *J. Neurosci.* **19**, 674–683 (1999).
- Swope, S. L., Moss, S. I., Raymond, L. A. & Haganir, R. L. Regulation of ligand gated ion channels by protein phosphorylation. *Adv. Second Messenger Phosphoprot. Res.* **33**, 49–78 (1999).
- Yan, Z. & Surmeier, D. J. D5 dopamine receptors enhance Zn²⁺ sensitive GABA_A currents in striatal cholinergic interneurons through a PKA/PP1 cascade. *Neuron* **19**, 1115–1126 (1997).
- Radnikow, G. & Misgeld, U. Dopamine D1 receptors facilitate GABA_A synaptic currents in the rat substantia nigra pars reticulata. *J. Neurosci.* **18**, 2009–2016 (1998).
- Moniyama, T. & Sim, J. A. Modulation of inhibitory transmission by dopamine in rat basal forebrain nuclei: activation of presynaptic D1 like dopaminergic receptors. *J. Neurosci.* **16**, 7502–7512 (1996).
- Brunig, I., Sommer, M., Hatt, H. & Bormann, J. Dopamine receptor subtypes modulate olfactory bulb γ -aminobutyric acid type A receptors. *Proc. Natl Acad. Sci. USA* **96**, 2456–2460 (1999).
- Bergson, C. et al. Regional cellular and subcellular variations in the distribution of D1 and D5 dopamine receptors in primate brain. *J. Neurosci.* **15**, 7821–7836 (1995).

- Buhl, E. H., Halasy, K. & Somogyi, P. Diverse sources of hippocampal unitary inhibitory postsynaptic potentials and the number of postsynaptic release sites. *Nature* **368**, 823–828 (1994).
- Nusser, Z., Roberts, J. D. B., Baude, A., Richards, J. G. & Somogyi, P. Immunocytochemical localization of $\alpha 1$ and $\beta 2/3$ subunits of the GABA_A receptor in relation to specific GABAergic synapses in the dentate gyrus. *Eur. J. Neurosci.* **7**, 630–636 (1995).
- Hall, R. A., Premont, R. T. & Lefkowitz, R. J. Heptahelical receptor signalling: beyond the G-protein paradigm. *J. Cell Biol.* **145**, 927–932 (1999).
- Sperk, G., Schwarzer, C., Tsunashima, K., Fuchs, K. & Sieghart, W. GABA_A receptor subunits in the rat hippocampus: I. Immunocytochemical distribution of 13 subunits. *Neuroscience* **80**, 987–1000 (1997).
- Essrich, C., Lorez, M., Berson, J. A., Fritschy, J.-M. & Luscher, B. Postsynaptic clustering of major GABA_A receptor subtypes requires the $\gamma 2$ subunit and gephyrin. *Nature Neurosci.* **1**, 563–571 (1998).
- Wang, H., Bedford, F. K., Brandon, N. J., Moss, S. J. & Olsen, R. W. GABA_A-receptor-associated protein links GABA_A receptors and the cytoskeleton. *Nature* **397**, 69–72 (1999).
- Gorrie, G. H. et al. Assembly of GABA_A receptors composed of $\alpha 1$ and $\beta 2$ subunits in both cultured neurons and fibroblasts. *J. Neurosci.* **17**, 6587–6596 (1997).
- Wan, Q. et al. Modulation of GABA_A receptor function by tyrosine phosphorylation of β subunits. *J. Neurosci.* **17**, 5062–5069 (1997).
- Luttrell, L. M., Ostrowski, J., Cotecchia, S., Kendall, H. & Lefkowitz, R. J. Antagonism of catecholamine receptor signaling by expression of cytoplasmic domains of the receptors. *Science* **259**, 1453–1557 (1993).
- Sugamori, K. S., Scheidele, M. A., Vernier, P. & Niznik, H. B. Dopamine D1B receptor chimeras reveal modulation of partial agonist activity by carboxyl terminal tail sequences. *J. Neurochem.* **71**, 2593–2599 (1998).
- McDonald, B. J. et al. Adjacent phosphorylation sites on GABA_A receptor β -subunits determine regulation by cAMP-dependent protein kinase. *Nature Neurosci.* **1**, 23–28 (1998).
- Hebert, T. E. & Bouvier, M. Structural and functional aspects of GPCR oligomerization. *Biochem. Cell Biol.* **76**, 1–11 (1998).
- Marshall, F. H., Jones, K. A., Kaupmann, K. & Bettler, B. GABA_A receptors—the first 7TM heterodimers. *Trends Pharmacol. Sci.* **10**, 369–399 (1999).
- Jordan, B. A. & Devi, L. A. G-protein coupled receptor heterodimerization modulates receptor function. *Nature* **399**, 697–700 (1999).
- Benes, F. M. The role of stress and dopamine-GABA interactions in the vulnerability for schizophrenia. *J. Psychiat. Res.* **31**, 257–275 (1997).
- Huntsman, M. M., Tran, B.-V., Potkin, S. G., Bunney, W. E. Jr & Jones, E. G. Altered ratios of alternatively spliced long and short $\gamma 2$ subunit mRNAs of the GABA_A receptor in prefrontal cortex of schizophrenics. *Proc. Natl Acad. Sci. USA* **95**, 15066–15071 (1998).
- Keverne, E. B. GABA-ergic neurons and the neurobiology of schizophrenia and other psychosis. *Brain Res. Bull.* **48**, 467–473 (1999).
- Okubo, Y. et al. Decreased prefrontal dopamine D1 receptors in schizophrenia revealed by PET. *Nature* **385**, 634–636 (1997).
- Goldman-Rakic, P. S. & Selemon, L. D. Functional and anatomical aspects of prefrontal pathology in schizophrenia. *Schiz. Bull.* **23**, 437–458 (1997).
- Moore, H. West, A. R. & Grace, A. A. The regulation of forebrain dopamine transmission. *Biol. Psychiat.* **46**, 40–55 (1999).
- Ng, G. Y. K. et al. Agonist induced desensitization of dopamine D1 receptor stimulated adenylyl cyclase activity is temporally and biochemically separated from D1 receptor internalization. *Proc. Natl Acad. Sci. USA* **92**, 10157–10161 (1995).
- Nusser, Z., Hajos, N., Somogyi, P. & Mody, I. Increased number of synaptic GABA_A receptors underlies potentiation at hippocampal inhibitory synapse. *Nature* **395**, 172–177 (1998).
- Wan, Q. et al. Recruitment of functional GABA_A receptors to postsynaptic domains by insulin. *Nature* **388**, 686–690 (1997).
- Kneussel, M. et al. Loss of postsynaptic GABA_A receptor clustering in gephyrin deficient mice. *J. Neurosci.* **19**, 9289–9297 (1999).
- Hall, R. A. et al. The $\beta 2$ adrenergic receptor interactions with the Na⁺/H⁺-exchanger regulatory factor to control Na⁺/H⁺ exchange. *Nature* **392**, 626–630 (1998).
- Yu, X.-M. & Salter, M. W. Gain control of NMDA receptor currents by intracellular sodium. *Nature* **396**, 469–474 (1998).

Acknowledgements

We thank J. Braunton, K. M. Zhu and H. Y. Man for technical assistance. This work was supported by grants from NIDA, the Ontario Mental Health Foundation (H.B.N.) and the MRC of Canada (H.B.N., X.M.Y., Y.T.W.). F.L. was supported by a C. Cleghorn Fellowship in Schizophrenia Research and is a fellow of the Canadian Psychiatric Research Foundation; Z.B.P. is a NARSAD Young Investigator. X.M.Y. is an MRC Scholar and Y.T.W. is a Research Scholar of the Heart and Stroke Foundation of Canada. H.B.N. was supported in part by the C.B. Ireland Endowed Fund for Psychiatric Research.

Correspondence and requests for materials should be addressed to Y.T.W. (e-mail: y.t.wang@utoronto.ca) or H.B.N. (hb.niznik@utoronto.ca).

Polarization rotation mechanism for ultrahigh electromechanical response in single-crystal piezoelectrics

Huaxiang Fu & Ronald E. Cohen

Carnegie Institution of Washington, 5251 Broad Branch Road, NW, Washington DC 20015, USA

Piezoelectric materials, which convert mechanical to electrical energy (and vice versa), are crucial in medical imaging, telecommunication and ultrasonic devices^{1,2}. A new generation of single-crystal materials³, such as $\text{Pb}(\text{Zn}_{1/3}\text{Nb}_{2/3})\text{O}_3\text{--PbTiO}_3$ (PZN–PT) and $\text{Pb}(\text{Mg}_{1/3}\text{Nb}_{2/3})\text{O}_3\text{--PbTiO}_3$ (PMN–PT), exhibit a piezoelectric effect that is ten times larger than conventional ceramics, and may revolutionize⁴ these applications. However, the mechanism underlying the ultrahigh performance of these new materials—and consequently the possibilities for further improvements—are not at present clear. Here we report a first-principles study of the ferroelectric perovskite, BaTiO_3 , which is similar⁵ to single-crystal PZN–PT but is a simpler system to analyse. We show that a large piezoelectric response can be driven by polarization rotation induced by an external electric field. Our computations suggest how to design materials with better performance, and may stimulate further interest in the fundamental theory of dielectric systems in finite electric fields.

A significant advance in seeking materials with high piezoelectric efficiency is marked by the growth of single-crystal solid solutions PZN–PT and PMN–PT. These materials exhibit—in contrast to the conventional piezoelectric ceramics—an ultrahigh piezoelectric coefficient d_{33} (the c -axis strain in response to a field along a cube axis c) of $\sim 2,500$ pC N^{−1}, a large strain level of 1.7%, and low hysteresis³. These solid solutions have a rhombohedral structure with polarization along the $\langle 111 \rangle$ direction. When an electric field is applied along the $\langle 001 \rangle$ direction, these materials respond mechanically in three stages as shown in Fig. 1a (to paraphrase the notation in ref. 3, we will call these stages A, B and C). Stages A and B are characterized by a dramatic increase of strain under a small field. Stage C has a very flat strain-versus-field slope, with almost no hysteresis. PZN–PT at stage C is indicated by X-ray diffraction to have an average tetragonal symmetry⁶.

PZN or PMN are relaxors (that is, dielectrics with a broad, frequency-dependent response as a function of temperature) with disorder on one site. Addition of a small amount of lead titanate (PT) to PZN or PMN turns them into normal ferroelectrics. The relaxor-to-ferroelectric phase transition is poorly understood. Both PZN–PT and PMN–PT are rhombohedral up to a critical PT content (9.5% in the former, and 35% in the latter) at which point they undergo a “morphotropic” phase transition and become tetragonal³. Giant piezoelectric response of d_{33} is observed in the rhombohedral phase of single crystals PZN–PT and PMN–PT.

The two most likely explanations for the giant piezoelectric response are as follows. The first is the reorientation of polar domains by the applied field. In this scheme, the fact that these materials are solid solutions with a relaxor end-member is crucial, but it is hard to understand why the materials must be single crystals. The second possibility was suggested³ to be polarization rotation, where the large response does not depend in an essential way on mesoscopic structure or ordering. While the field-induced polarization rotation was shown⁷ to indeed occur, there is, however,

little understanding of why this rotation can give rise to the large piezo-response.

We now show why this polarization rotation can drive a giant piezoelectric response, revealing the mechanism at an atomistic level. The theoretical piezo-response that we derive (Fig. 1b) indeed shows a large piezoelectric coefficient at small field, as observed³ in the single-crystal materials. The intuitive explanation for this mechanism is outlined as follows. PbTiO_3 is a tetragonal ferroelectric with a large strain⁸ (6%); if there were a rhombohedral PT phase, one could expect to drive a colossal mechanical strain with application of an electric field. Tetragonal BaTiO_3 , which does have a rhombohedral ground state, in contrast has a much smaller strain (1%). The new piezoelectrics are, owing to their compositions, like “rhombohedral” PT, which allows the large observed response.

Here we study ferroelectric perovskite BaTiO_3 , which is tractable, rather than solid solutions. A large piezoelectric enhancement induced by a $\langle 001 \rangle$ field has been reported⁵ in rhombohedral BaTiO_3 as in PZN–PT, and BaTiO_3 is simple in structure, which allows us to determine the essential physics. The complex mesoscopic ordering⁹ in PZN–PT or PMN–PT greatly complicates their study from first principles⁸, but we propose that the same behaviour demonstrated here for BaTiO_3 should also be true for the more complex materials.

The phase stability under electric field $\mathbf{E}$ is studied using free energy $F = U - \mathbf{P} \cdot \mathbf{E}$, where U is the internal energy under zero field and $\mathbf{P}$ is the polarization per unit cell. Instead of directly minimizing the free energy for a given field, which is not possible, we select different polarization directions and calculate their internal energies using the first-principles linearized augmented plane wave (LAPW) method^{10,11} with the local density-functional approximation (LDA). The polarization direction is controlled by constraining the moving direction of the Ti atom, and other degrees of freedom (including both atomic positions and cell shape) are optimized (while keeping the cell volume fixed at the experimental value $64.2 \text{ \AA}^3$). The final

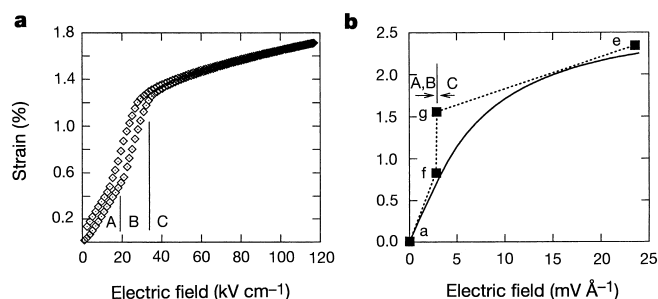


Figure 1 Strain responses under electric field. **a**, Experimental measurements for PZN–8%PT (ref. 3). **b**, Theoretical results for BaTiO_3 . The square symbols (connected by a dotted line as a guide for eyes) are directly obtained from first-principles calculations. The solid curve is obtained by (1) fitting the internal energies and strains in Table 1 as functions of polarization angle using, respectively, crystal harmonics with cubic symmetry to angular momentum $l = 8$ and crystal harmonics with tetragonal symmetry to $l = 6$; and (2) minimizing the free energy with respect to the polarization for a given field. (The fitting expressions and coefficients are available at <http://www.gi.ciw.edu/~cohen/research/btostrain/btospherfit.html>). Both the first-principles and the analytically-obtained results reveal a large strain-versus-field slope at small fields. The larger fitting error near polarization g may be due to its large deviation of polarization magnitude from the average value. Our calculations show that an electric field $E = 22 \text{ mV \AA}^{-1}$ ($1 \text{ mV \AA}^{-1} = 10^2 \text{ kV cm}^{-1}$) is needed to change BaTiO_3 from the rhombohedral phase to the perfectly tetragonal, which is in accord with $E = 25 \text{ mV \AA}^{-1}$ obtained⁷ using the effective hamiltonian. However, instead of a sublinear curve with a small slope⁷, we find here very different strain-versus-field regimes at low and high fields. This discrepancy probably results from the fact that we use first-principles methods directly rather than a model hamiltonian: the latter may miss the small-energy-scale behaviour in describing the detailed strain versus finite-field response. We also note, however, that our calculation is for zero temperature and wavevector $q = 0$, so details will change for finite temperatures.

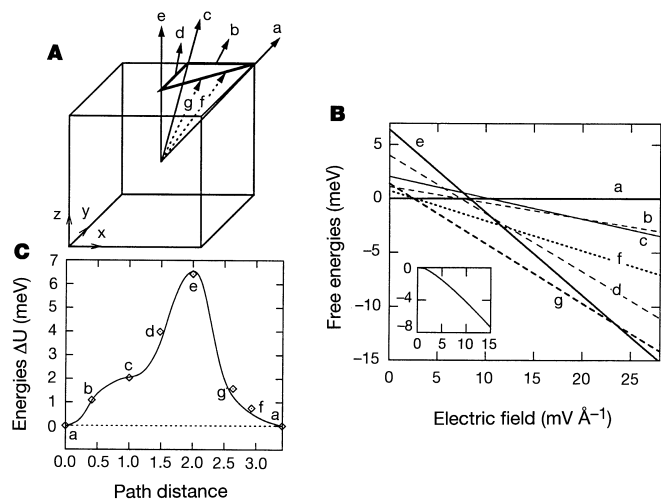


Figure 2 Free and internal energies. **A**, Schematic illustration of the polarization directions considered in our calculations. The lowest free energy path is found along the path $a \rightarrow f \rightarrow g \rightarrow e$. **B**, Free energies as a function of field strength for different polarization directions which are given as labels near each line. The free energy of polarization direction a is taken as zero reference for clarity of display. The inset shows the free energy difference (in meV) of the rotation along $a \rightarrow f \rightarrow g \rightarrow e$ relative to the rotation along $a \rightarrow b \rightarrow c \rightarrow d \rightarrow e$ as a function of field (in $\text{mV } \text{\AA}^{-1}$). **C**, The internal energies $\Delta U = U - U_{\text{rhomb}}$ relative to the rhombohedral phase (symbols, LAPW results; line, a guide to the eyes) along the closed path $a \rightarrow b \rightarrow c \rightarrow d \rightarrow e \rightarrow g \rightarrow f \rightarrow a$, where the horizontal axis is the path distance along the heavy-line triangle in panel **A**.

polarization $\mathbf{P}$ is computed from the optimized geometric structure using the Berry's phase approach^{12,13} (see ref. 14 for its implementation in LAPW). Note that all of our computations are in fact at zero field, though we are extracting finite-field results using the above approach. At present there is no fundamental theory¹⁵ for ferroelectric systems under finite electric fields because of the unknown field dependence of the energy functional, the non-analytical form of polarization with respect to electronic wavefunctions, and the loss of periodicity in the potential.

We consider the polarization directions shown in Fig. 2A. The calculated internal energies and polarization magnitudes are given in Table 1, which shows the increasing phase stability in the order of tetragonal $\rightarrow$ orthorhombic $\rightarrow$ rhombohedral under zero field. Our calculated spontaneous polarizations per unit volume for these three phases are, respectively, 0.44, 0.43 and 0.45 C m^{-2} , agreeing with pseudopotential results¹⁶ (0.36, 0.43 and 0.43 C m^{-2}). We note however that the calculated strain ϵ_{33} of the tetragonal phase (2%) is too large compared with the experimental value¹⁷ (1%), similar to what is found for PbTiO_3 using LDA theory¹⁴. We find for BaTiO_3 that the generalized gradient approximation (GGA) gives $\epsilon_{33} = 1.07\%$, in much better agreement with experiment. However, our main conclusion, which is mainly in the vicinity of rhombohedral phase (see below), should not be affected, and we use LDA because of its computational efficiency.

The free energies as a function of field strength are shown in Fig. 2B. Using Fig. 2B and Table 1, the strain achievable for a given field can be obtained; such strains are shown in Fig. 1b for polarization rotation along directions $a \rightarrow f \rightarrow g \rightarrow e$ (these directions are defined in Fig. 2A). In order to have a continuous strain-field curve for any given field, we fit the obtained internal energies and strains as functions of polarization angle using the symmetrized lattice harmonic expansions, and the polarization direction for a given field is obtained by minimizing the analytically obtained free energy. The minimum energy path is found to be directly from the $\langle 111 \rangle$ to $\langle 001 \rangle$ directions (that is, the path $a \rightarrow f \rightarrow g \rightarrow e$ in Fig. 2A). We

see in Fig. 1b that the strain increases rapidly in the low-field region, and a small electric field will induce a large strain level (thus a large piezoelectric coefficient). We also find that, if the rotation path $a \rightarrow b \rightarrow c \rightarrow d \rightarrow e$ is possible, the piezo-coefficient d_{33} of this path will be 5 times as small as that of the path $a \rightarrow f \rightarrow g \rightarrow e$ in the low-field region.

To analyse what causes the large piezo-response along the rotation path $a \rightarrow f \rightarrow g$, the internal energies for the considered polarizations are shown in Fig. 2C, where we see a rather flat internal energy surface along the path $a \rightarrow f \rightarrow g$. Furthermore, Table 1 shows that along this path there is a large increase of c -axis polarization P_z (thus a large coupling between the polarization and the electric field). The flat internal energy surface and the large c -axis polarization variation facilitate the polarization rotation, and cause the peculiarly large piezo-response along this path.

Our theory provides explanations for the experimental observations^{3,5}. In the low-field region of stage A, a piezoelectric coefficient d_{33} of 350 pC N^{-1} was reported⁵ for BaTiO_3 at a temperature of -100°C . Our theoretical finite-difference piezoelectric coefficient is 293 pC N^{-1} , obtained from polarization direction f . Both values are larger than the experimental d_{33} of the tetragonal phase^{5,18} ($\sim 100 \text{ pC N}^{-1}$). In the field region corresponding to the transition from stage B to stage C, the experimental data for BaTiO_3 are not available at low temperature because a large field is needed. In this region, our theoretical strain behaviour mimics the experimental result³ in PZN-PT (see Fig. 1); that is, a marked increase of strain under a small field at stages A and B, and then a very slow strain increase at stage C. When the polarization changes into direction g (that is, near the end of stage B), the strain level reaches up to 68% of the full strain of the tetragonal phase according to our calculation. The material-specific difference between BaTiO_3 and PZN-PT probably acts mainly to scale the magnitude of transition field: a very large field ($\sim 300 \text{ kV cm}^{-1}$) is needed to transform BaTiO_3 from stage B to stage C (Fig. 1b), whereas a much smaller field ($\sim 40 \text{ kV cm}^{-1}$) was shown³ to be able to trigger this transition in PZN-8%PT. This explains why PZN-PT is a highly efficient piezoelectric and BaTiO_3 is not.

Our computations are for zero temperature and for a single-domain crystal. At finite temperature BaTiO_3 displays both order/disorder and displacive character. In the ideal order/disorder case, domains have polarizations along four $(\pm 1 \pm 1)$ directions and will respond equivalently to the field along the $\langle 001 \rangle$ direction: thus our zero-temperature computation will also be appropriate for the pure field effect. Note that the zero-field thermally excited tetragonal phase of BaTiO_3 differs from the field-induced tetragonal phase in that the former is partially disordered whereas the latter is ordered. Temperature may also alter the relative free energies; for instance, near the orthorhombic-tetragonal transition temperature a different piezo-response starting from the orthorhombic phase would be

Table 1 Internal energies, polarization magnitudes, and strains.

Polarization	ΔU	P_x	P_y	P_z	ϵ_{33}
a (rhomb.)	0.00	1.98	1.98	1.98	0.00
b	1.09	1.33	2.26	2.26	0.45
c (orth.)	2.04	0.00	2.36	2.36	0.55
d	3.99	0.00	1.55	3.00	1.64
e (tetr.)	6.42	0.00	0.00	3.44	2.34
f	0.78	1.65	1.65	2.51	0.84
g	1.58	1.36	1.36	3.03	1.56

Energy difference $\Delta U = U - U_{\text{rhomb}}$ (in meV) relative to the rhombohedral phase, polarization magnitudes per bulk cell along three directions P_x , P_y and P_z (in units of $e \times \text{bohr}$), and strain ϵ_{33} ($\epsilon_{33} = c - a_0/a_0$, where c and a_0 are the cell lengths of strained and unstrained systems, respectively) in units of %, for different polarization directions. Calculations for internal energy U are converged; we used $RK_{\text{max}} = 8$ and a $6 \times 6 \times 6$ special K -point mesh. The energy minimization is guided by atomic force and finite-difference stress; we find that the c/a ratio is the dominating effect in cell-shape minimization for monoclinic structure. Polarization $\mathbf{P}$ is, to the second order, related to the strain $\{\epsilon_{ij}\}$ and internal displacements $\mathbf{u}^i$ of atom α by¹³ $P_k = e_{ijk}^0 \epsilon_{ij} + \Sigma_\alpha Z_\alpha^k u_\alpha^i + \chi E_k$, where the three polarization terms are from the homogeneous strain, the atomic displacement carrying effective charge Z^k , and the electronic susceptibility χ . The third term, which will change only slightly by the polarization rotation, is neglected here. The effective charges Z^k are greatly enhanced in ferroelectrics^{12,16}; through these charges, the piezoelectric effect arises primarily from the coupling of the internal displacements with strain.

expected, and finite-temperature simulations⁷ would be necessary.

PZN–PT at stage C has been shown by X-ray diffraction to be tetragonal⁶, which, as discussed above, can be disordered (an average structure) or ordered, depending on the applied field and temperature. Our calculation suggests that stage C could be just a part of the rotation path, and the tetragonal diffraction may be due to some disorder. The weaker piezo-response at stage C results from the large internal energy barrier (Fig. 2C) and the little polarization variation (Table 1) between polarizations g and e. Our calculation does not convincingly reproduce the observed³ slight difference between stages A and B. This difference could come from the fact that stage A probably involves domain alignment, and stage B may involve two-phase coexistence.

Our conclusion that the polarization rotation alone can result in the giant piezoelectric response is important: efforts to design better materials can focus on the intrinsic structure rather than the extrinsic nanodomains. Our calculations further suggest that, in addition to a rhombohedral phase at zero field and a high strain level in the tetragonal phase, good materials need to have: (1) a flat energy surface (that is, soft force constants for ferroelectric displacements) near the rhombohedral phase, so that a small field will cause a large change of polarization angle; (2) large effective charges (thus a large *c*-axis polarization variation to be coupled with electric field during polarization rotation); and (3) strong coupling between the internal degrees of freedom and strain. Polarization rotation along the minimum free energy path minimizes the repulsion between Ti and O in the base of the (oxygen) octahedron, which favours the flat energy surface. Simultaneously, this rotation path increases the repulsion between Ti and O at the top of the octahedron. A tetragonal strain can ease the latter repulsion, which causes the strong coupling of internal freedoms with strain. From our results (Table 1) we see that the flat energy surface near the rhombohedral phase may be the most important factor. In single crystals the local polarizations will be rotated coherently from $\langle 111 \rangle$ to $\langle 001 \rangle$ directions, while the effect will largely cancel in a randomly oriented ceramic.

We note that the polarization rotation effect under finite fields may not manifest itself in the region of infinitesimal field. As there is at present no density functional theory for systems under finite electric fields, the work that we report here may stimulate an interest in this area. □

Received 16 July; accepted 19 November 1999.

1. Lines, M. E. & Glass, A. M. Principles and applications of ferroelectrics and related materials (Clarendon, Oxford, 1979).
2. Uchino, K. Piezoelectric actuators and ultrasonic motors (Kluwer Academic, Boston, 1996).
3. Park, S. -E. & Shrout, T. R. Ultrahigh strain and piezoelectric behavior in relaxor based ferroelectric single crystal. *J. Appl. Phys.* **82**, 1804–1811 (1997).
4. Service, R. F. Shape-changing crystals get shifter. *Science* **275**, 1878 (1997).
5. Park, S. -E. *et al.* Crystallographic engineering in high-performance piezoelectric crystals. *J. Appl. Phys.* (submitted).
6. Durbin, M. K., Jacobs, E. W., Hicks, J. C. & Park, S. -E. In situ x-ray diffraction study of an electric field induced phase transition in the single crystal relaxor ferroelectric, 92% Pb(Zn_{1/3}Nb_{2/3})O₃-8%PbTiO₃. *Appl. Phys. Lett.* **74**, 2848–2850 (1999).
7. Garcia, A. & Vanderbilt, D. Electromechanical behavior of BaTiO₃ from first-principles. *Appl. Phys. Lett.* **72**, 2981–2983 (1998).
8. Cohen, R. E. Origin of ferroelectricity in oxide ferroelectrics and the difference in ferroelectric behavior of BaTiO₃ and PbTiO₃. *Nature* **358**, 136–138 (1992).
9. Egami, T., Dmowski, W., Akbas, M. & Davies, P. K. in *First-principles Calculations for Ferroelectrics: Fifth Williamsburg Workshop* (ed. Cohen R. E.) 1–10 (AIP, Woodbury, New York, 1998).
10. Wei, S. H. & Krakauer, H. Local-density-functional calculation of the pressure-induced metallization BaSe and BaTe. *Phys. Rev. Lett.* **55**, 1200–1203 (1985).
11. Singh, D. J. *Planewaves, Pseudopotentials and the LAPW Method* (Kluwer, Boston, 1994).
12. King-Smith, R. D. & Vanderbilt, D. Theory of polarization of crystalline solids. *Phys. Rev. B* **47**, 1651–1654 (1993).
13. Resta, R. Macroscopic polarization in crystalline dielectrics: the geometric phase approach. *Rev. Mod. Phys.* **66**, 899–915 (1994).
14. Saghi-Szabo, G., Cohen, R. E. & Krakauer, H. First-principles study of piezoelectricity in PbTiO₃. *Phys. Rev. Lett.* **80**, 4321–4324 (1998).
15. Martin, R. M. & Ortiz, G. Functional theory of extended Coulomb systems. *Phys. Rev. B* **56**, 1124–1140 (1997).
16. Ghosez, Ph., Michenaud, J.-P. & Gonze, X. The physics of dynamical atomic charges: The case of ABO₃ compounds. *Phys. Rev. B* **58**, 6224–6239 (1998).

17. Mitsui, T. *Ferroelectrics and Related Substances* (eds K. H. Hellwege & O. Madelung) Landolt-Bornstein, New Series Group III, Vol. 16, Pt a (Springer, Berlin, 1981).
18. Zgonik, M. *et al.* Dielectric, elastic, piezoelectric, electro-optic, and elasto-optic tensors of BaTiO₃ crystals. *Phys. Rev. B* **50**, 5941–5949 (1994).

Acknowledgements

We thank H. Krakauer and D. Singh for sharing their LAPW codes and discussions; S.E. Park for sending us the experimental data in Fig. 1a and ref. 5; and D. Vanderbilt, S.E. Park, T. Shrout, W. Smith and K. Rabe for discussions. This work was supported by the Office of Naval Research. Computations were performed on the CRAY SV1 supported by the NSF and the Keck Foundation.

Correspondence and requests for materials should be addressed to R.E.C. (e-mail: cohen@gl.ciw.edu).

Dynamics of supercooled water in confined geometry

R. Bergman* & J. Swenson†

* Department of Experimental Physics, † Department of Applied Physics, Chalmers University of Technology, SE-412 96 Göteborg, Sweden

As with most liquids, it is possible to supercool^{1–4} water; this generally involves cooling the liquid below its melting temperature (avoiding crystallization) until it eventually forms a glass. The viscosity and related relaxation times (τ) of glass-forming liquids typically show non-Arrhenius temperature (T) dependencies. Liquids with highly non-Arrhenius behaviour in the supercooled region are termed ‘fragile’. In contrast, liquids whose behaviour is close to the Arrhenius law ($\ln \tau \propto 1/T$) are termed ‘strong’ (ref. 5). A unique ‘fragile–strong’ transition around 228 K has been proposed⁶ for supercooled water; however, experimental studies of bulk supercooled water in this temperature range are generally hampered because crystallization occurs. Here we use broad-band dielectric spectroscopy to study the relaxation dynamics of supercooled water in a wide temperature range, including the usually inaccessible temperature region. This is possible because the supercooled water is held within a layered vermiculite clay—the geometrical confinement and presence of intercalated sodium ions prevent⁷ most of the water from crystallizing. We find a relaxational process with an Arrhenius temperature dependence, consistent with the proposed strong nature of deeply supercooled bulk water. Because water that is less supercooled has been established⁶ as highly fragile, our results support the existence of a fragile–strong transition.

Although water has been the topic of considerable research, its peculiar physical properties are not well understood—it remains in many ways a mysterious material⁸. Below the melting point at 273 K it is possible to supercool normal bulk water down to ~ 235 K where it crystallizes. It is also possible to rapidly quench small quantities of water to temperatures below 100 K to form amorphous solid water, that is, glassy water. When heating glassy water it exhibits a glass–liquid transition at $T_g \approx 124$ –136 K (refs 9, 10) and then, because of the increased mobility of the molecules, crystallizes at 150 K (ref. 2). There is thus an experimentally inaccessible temperature region for bulk supercooled water between roughly 150 and 235 K. It has been shown that the temperature dependence of the viscosity and the accompanying main relaxation time τ , above this temperature region, follows a power law that diverges at a critical temperature $T_s \approx 228$ K (refs 11, 12) in the inaccessible temperature region.

The most fragile liquids are found among weakly interacting

ionic and van der Waals systems, such as salol (phenyl salicylate), while strong supercooled liquids typically are network glass-formers, such as SiO_2 and GeO_2 . Water above the inaccessible temperature region has been suggested to be an extremely fragile liquid⁶. However, there must be a change in the temperature dependence of the relaxation time (or viscosity) in this temperature region, as T_g , which corresponds to a finite relaxation time of about 100 s, is at a considerably lower temperature than T_g . The idea of a fragile–strong transition is also supported by thermodynamic data⁶.

Confined water is of importance in biological systems and may also provide a means of entering into the inaccessible temperature region for bulk supercooled water. In certain geometrical confinements the water molecules are unable to form a crystalline structure; the water thus remains fluid and, like any other supercooled liquid, becomes increasingly viscous as the temperature approaches T_g . The dynamics will be affected by the confinement; the question is how it will differ from bulk dynamics. As a result of a continuing debate about the existence of a heterogeneous dynamic structure with regions of cooperative motions in glass-forming liquids¹³, there have been several studies of glass-formers in confined geometry¹⁴. It has been shown that confinement decreases the α -relaxation time, that is lowering the T_g (refs 15, 16). However, the confinement effects are less pronounced in two-dimensional confinements than in three-dimensional confinements such as pores¹⁵. Here we choose to study the dynamics of water confined in only one direction, that is, the water molecules are restricted to move in a two-dimensional structure.

Vermiculite clays are unique in the sense that they are able to provide macroscopic samples of two-dimensional stack geometries with extremely thin and well-defined thickness of water layers between rigid clay platelets¹⁷ (that is, one-dimensional confinement). Neutron spectroscopy has shown^{18–21} that the intercalated water of vermiculite clays has similar dynamics in a direction parallel to the clay layers to that of room temperature and slightly supercooled²¹ bulk water. Earlier work^{18,19} gave a too-strong dependence of the thickness of the water layer as a result of insufficient resolution and the analysis method used²⁰. The bulk-like dynamics may be explained by results from a neutron diffraction study of Na-vermiculite¹⁷ which has shown that there are water molecules that

interact only with other water molecules for a distance between clay platelets corresponding to two layers of water molecules. The thickness of the water layer in these clays can be regulated very accurately in three discrete steps corresponding to zero, one and two water layers between platelets simply by varying the relative humidity. In the one-water-layer phase (with a spacing of 11.78 Å between clay platelets) all the hydrogen atoms are bound directly to the adjacent platelets¹⁷. Hydrogen bonds between the water molecules occur in the two-water-layer phase (14.96 Å between platelets), where only about half of the water molecules form a hydrogen bond to one of the clay surfaces¹⁷. Differential scanning calorimetry⁷ shows that approximately 24% of the interlayer water in the two-water-layer system crystallizes at 235 K whereas the remaining water remains in a supercooled liquid state. We believe that the presence of sodium ions introduce the disorder necessary to prevent total crystallization of the interlayer water. Nucleation probably occurs either in voids formed by structural defects of the clay platelets or in small regions relatively far from neighbouring sodium ions where the surrounding disorder interrupts further nucleation. The structure of this mixture of crystalline and supercooled water will soon be investigated by neutron diffraction.

To study the dynamics of the water confined in these Na-vermiculite clays we performed broad-band dielectric spectroscopy⁷ (Novocontrol, 10^{-2} – 10^7 Hz) and quasi-elastic neutron scattering²¹ in controlled humidity and temperature. For reference we also measured the dielectric response of bulk ice. As expected, the dielectric spectra of the phases (corresponding to zero, one and two water layers between platelets) show large differences (Fig. 1). The permittivity data for the system saturated with water (two-water-layers) show a pronounced relaxation peak moving towards higher frequencies as temperature is increased (Fig. 2). This strong peak is not observed in the system with one water layer and is therefore attributed to the dipolar relaxation of the water molecules that are only interacting with other water molecules. We can follow the peak from approximately 130 K where it enters our dynamical window at the low-frequency side to roughly 215 K where it is swamped by a low frequency dispersion.

Between 160 K and 215 K the relaxation peak is symmetric and can be described by the empirical Cole–Cole equation. We note that very recent dielectric²² and Raman²³ measurements at higher frequencies on bulk water at higher temperatures have also been described using a Cole–Cole equation. For the high temperature spectra a power law was used in the curve-fitting routine for the increasingly dominant low-frequency dispersion. The exponent of the power-law was found to be about 0.5 (less than unity) in the real part as well as in the imaginary part of the permittivity. This suggests

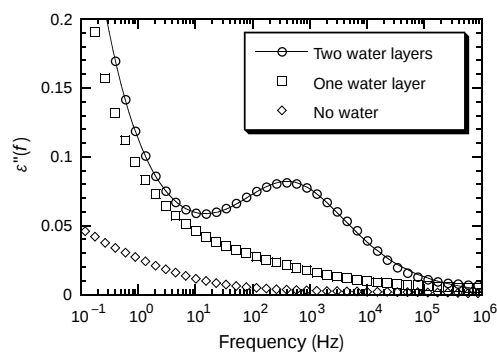


Figure 1 Dielectric spectra. Frequency (f) dependence of the imaginary part $\epsilon''(f)$ of the dielectric relaxation $\epsilon^*(f)$ of the vermiculite clays at 185 K is shown. The low-frequency behaviour of the one- and two-water-layer systems is very similar. At higher frequencies, a pronounced peak in the two-water-layer system is absent in the one-water-layer system. The line through the data points of the two-water-layer spectrum is a least-square fit to a superposition of a power law, proportional to f^{-n} , where $n \approx 0.5$, and the imaginary part of a Cole–Cole function: $\epsilon''(f) = \epsilon_\infty + (\epsilon_s - \epsilon_\infty) \frac{1}{1 + (2\pi f \tau)^n}$. Here ϵ_s and ϵ_∞ are the static and high-frequency limiting values of the dielectric constant, respectively, and τ is the relaxation time. The width parameter α was found to be ~ 0.55 at 185 K. Although α is roughly constant below this temperature there is a slight increase in this value, such that the response approaches a Debye behaviour ($\alpha = 1$) as temperature is increased.

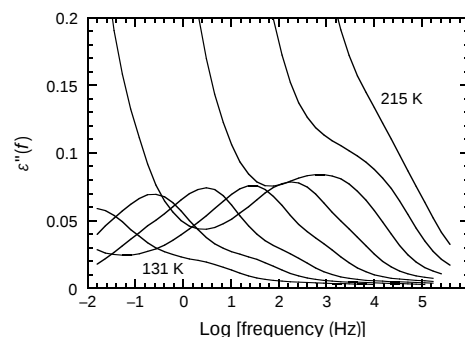


Figure 2 $\epsilon''(f)$ of the two-water-layer system at some representative temperatures. Spectra from left to right were recorded at 131 K and then at every twelfth degree until 215 K. We note that the spectra at the lowest temperatures exhibit a bimodal peak, and the higher-temperature spectra are essentially monomodal. However, at these high temperatures the relaxation spectra are strongly affected by a power-law dispersion.

that the process is not due to normal d.c. conductivity, which would not contribute to the real part. It is also difficult to envisage a d.c. conductivity perpendicular to the platelets. Rather, we attribute this process to the so-called anomalous low-frequency dispersion (ALFD)²⁴ which has also been observed in other, structurally similar, layered materials²⁵.

At lower temperatures the relaxation peak is clearly bimodal (Fig. 2), suggesting that the water at these temperatures exists in two different states. At low temperatures a very small third peak can be detected at higher frequencies⁷. To describe the data at lower temperatures it is therefore necessary to use a superposition of at least two Cole–Cole functions. The second peak is not as broad as the main one, and due to its small amplitude could not be observed above ~ 160 K.

In Fig. 3 we show the two main relaxation times obtained from the fits, together with data from the literature. The relaxation time for the second peak coincides with that observed for ice and this process is therefore attributed to crystalline water molecules. The slowest relaxation, not seen in the ice sample or in the zero- and one-water-layer clay systems, has an Arrhenius temperature dependence and a relaxation time of 100 s ($\sim \tau(T_g)$) at ~ 127 K which is close to the reported T_g of water (124–136 K) (refs 9, 10). We therefore attribute this process to the reorientation of supercooled water molecules that have more bulk-like dynamical properties in our system. Reassuringly, the temperature dependence of the relaxation times for both observed relaxations extrapolates to physically realistic values of $\sim 10^{-14}$ s at infinite temperatures, that is, times corresponding to quasi-lattice vibrations. Note also, in Fig. 3, that the high-temperature dielectric²⁶ and neutron-

scattering²⁰ data of bulk water coincides with neutron spectroscopy data on the present and similar systems^{18–21}, even in the moderately supercooled state²¹. Thus, as the relaxation times at both T_g and room temperature are consistent with bulk values, we believe that at least some of the confined water molecules show dynamical similarities to bulk water. The Arrhenius behaviour at low temperatures implies that this confined water is a strong glass-former at T_g .

The onset of strong behaviour at around T_s has been suggested to be due to an end point in the hydrogen-bond network formation process^{6,12}. If so, the observed main broad relaxation peak would be caused by the reorientation of water molecules in an OH-bonded network. This is very similar to the situation in ice²⁷. However, in the present case the network is disordered and mechanically soft, giving a broad distribution of activation energies that in turn implies a broad distribution of relaxation times. This explains the low value of the Cole–Cole parameter ($\alpha \approx 0.55$).

One indication of the fragility is m , the slope of the logarithm of the relaxation time versus T_g/T at T_g (ref. 28). Our relaxation-time data of the main and slowest process gives a very low value of $m \approx 14$ which is even lower than for SiO_2 and GeO_2 , both archetypical strong glass-formers²⁸ with $m \approx 20$. Thus, if we attribute the main peak to the α -relaxation, this implies that our confined water, close to its T_g , is one of the strongest glass-formers studied. This is remarkable as bulk water at higher temperatures is considered to be an extremely fragile liquid⁶. □

Received 21 June; accepted 22 November 1999.

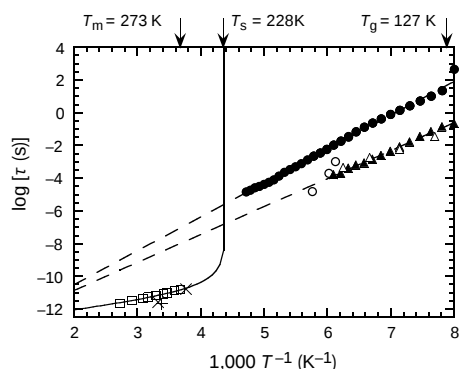


Figure 3 Temperature dependence of the relaxation times obtained from fits of the spectra from the two-water-layer clay. The filled circles denote the relaxation times for the main relaxation and the filled triangles show the relaxation time of the weaker second process observed at low temperatures. The open triangles coinciding with the second process are from a bulk water (ice) sample. The second process is therefore attributed to crystalline water in, or on the surface of, the clay sample. The dashed straight lines are fits to the Arrhenius equation, $\tau = \tau_0 \exp(E/RT)$. Also included are new literature data (open squares) of the dielectric relaxation time of bulk water at higher temperatures²⁶. The full line through the data is a power-law (the exponent is 1.55) diverging at $T_s = 228$ K (ref. 11). The residence times ($\approx \tau$) of the neutron-scattering data of our system (slanted crosses) were obtained from using a translational jump-diffusion model²⁰ and coincides with the neutron-scattering (upright crosses) (ref. 20) and the dielectric data of bulk water. The data points (open circles) at lower temperatures are from dielectric measurements on water sequestered in pores of a polymer matrix²⁹. These data suggest a more fragile behaviour also close to T_g . However, the sequestered water is confined in all three dimensions, whereas our system is confined to the two-dimensional interplatelet planes, that is, the structure only exhibits one-dimensional confinement. Three-dimensional confinement has been shown to speed up the α -relaxation of glass-formers more than two-dimensional confinement does¹⁵. T_m is the melting temperature; T_s is the critical divergence temperature; T_g is the glass–liquid transition temperature.

1. Tanaka, H. A self-consistent phase diagram for supercooled water. *Nature* **380**, 328–330 (1996).
2. Mishima, O. & Stanley, H. E. The relationship between liquid, supercooled and glassy water. *Nature* **396**, 329–335 (1998).
3. Angell, C. A. Supercooled water: approaching the limits. *Nature* **331**, 206–207 (1988).
4. Smith, R. S. & Kay, B. D. The existence of supercooled liquid water at 150 K. *Nature* **398**, 788–791 (1999).
5. Angell, C. A. Relaxations in liquids, polymers and plastic crystals—strong/fragile patterns and problems. *J. Non-Cryst. Solids* **131–133**, 13–31 (1991).
6. Ito, K., Moynihan, C. T. & Angell, C. A. Thermodynamic determination of fragility in liquids and a fragile-to-strong liquid transition in water. *Nature* **398**, 492–495 (1999).
7. Bergman, R., Swenson, J. & Björjesson, L. Dielectric study of supercooled 2D-water in a vermiculite clay. *J. Chem. Phys.* (submitted).
8. Stanley, H. E. Unsolved mysteries of water in its liquid and glass states. *Mater. Res. Bull.* **24**, 22–30 (1999).
9. Handa, Y. P. & King, D. D. Heat capacity and glass transition behavior of amorphous ice. *J. Phys. Chem.* **92**, 3323–3325 (1988).
10. Johari, G. P., Hallbrucker, A. & Mayer, E. The glass-liquid transition of hyperquenched water. *Nature* **330**, 552–553 (1987).
11. Angell, C. A. in *Water: A Comprehensive Treatise* (ed. Franks, F.) 1–81 (1982).
12. Speedy, R. J. & Angell, C. A. Isothermal compressibility of supercooled water and evidence for a thermodynamic singularity at -45°C . *J. Chem. Phys.* **65**, 851–858 (1976).
13. Sillescu, H. Heterogeneity at the glass transition: a review. *J. Non-Cryst. Solids* **243**, 81–108 (1999).
14. Ngai, K. L., Riande, E. & Wright, G. B. (eds) Proceedings of the Third International Discussion Meeting on Relaxations in Complex Systems. *J. Non-Cryst. Solids* **235–237**, (1998).
15. Barut, G., Pissis, P., Pelster, R. & Nimtz, G. Glass transition in liquids: Two versus three-dimensional confinement. *Phys. Rev. Lett.* **80**, 3543–3546 (1998).
16. Arndt, M., Stannarius, R., Groothues, H., Hempel, E. & Kremer, F. Length scale of cooperativity in the dynamic glass transition. *Phys. Rev. Lett.* **79**, 2077–2080 (1997).
17. Skipper, N. T., Soper, A. K. & McConnell, D. C. The structure of interlayer water in vermiculite. *J. Chem. Phys.* **94**, 5751–5760 (1991).
18. Olejnic, S. & White, J. W. Thin layers of water in vermiculites and monmorillonites-modification of water diffusion. *Nature* **236**, 15–16 (1972).
19. Hunter, R. J., Stirling, G. C. & White, J. W. Water dynamics in clays by neutron spectroscopy. *Nature* **230**, 192–194 (1971).
20. Tuck, J. J., Hall, P. L., Hayes, M. H. B., Ross, D. K. & Poinson, C. Quasi-elastic neutron-scattering studies of the dynamics of intercalated molecules in charge-deficient layer silicates. *J. Chem. Soc., Faraday Trans.* **80**, 309–324 (1984).
21. Swenson, J., Bergman, R. & Howells, W. S. Quasi-elastic neutron scattering of 2D-water in a vermiculite clay. *J. Chem. Phys.* (submitted).
22. Okada, K., Yao, M., Hiejima, Y., Kohno, H. & Kajihara, Y. Dielectric relaxation of water and heavy water in the whole fluid phase. *J. Chem. Phys.* **110**, 3026–3036 (1999).
23. Mizoguchi, K., Ujike, T. & Tominaga, Y. Dynamical structure of water in NaCl aqueous solution. *J. Chem. Phys.* **109**, 1867–1872 (1998).
24. Jonscher, A. K. Low-frequency dispersion in carrier-dominated dielectrics. *Phil. Mag. B* **38**, 587–601 (1978).
25. Jeevanandam, P. & Vasudevan, S. Anomalous low frequency dispersion and dielectric relaxation in the layered intercalated compounds $\text{Cd}_{0.75}\text{PS}_{3.25}\text{A}_{0.25}(\text{H}_2\text{O})$ [$\text{A}=\text{K}, \text{Cs}$]. *J. Chem. Phys.* **108**, 1206–1215 (1998).
26. Ronne, C., Åstrand, P.-O. & Keiding, S. R. THz spectroscopy of liquid H_2O and D_2O . *Phys. Rev. Lett.* **82**, 2888–2891 (1999).

27. Bjerrum, N. Structure and properties of ice. *Science* **115**, 385–390 (1952).
 28. Böhrer, R., Ngai, K. L., Angell, C. A. & Plazek, D. J. Nonexponential relaxations in strong and fragile glass formers. *J. Chem. Phys.* **99**, 4201–4209 (1993).
 29. Johari, G. P. Water's character from dielectric relaxation above its T_g . *J. Chem. Phys.* **105**, 7079–7082 (1996).

Acknowledgements

We thank C. A. Angell and L. Börjesson for discussions; M. V. Smalley for providing the vermiculite clay samples; and P. Jacobsson for discussions and the use of dielectric spectrometer equipment.

Correspondence and requests for materials should be addressed to R.B. (e-mail: f5xrb@fy.chalmers.se).

Molecular imprinting of bulk, microporous silica

Alexander Katz[†] & Mark E. Davis^{*}

^{*} Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA

Molecular imprinting aims to create solid materials containing chemical functionalities that are spatially organized by covalent¹ or non-covalent² interactions with imprint (or template) molecules during the synthesis process. Subsequent removal of the imprint molecules leaves behind designed sites for the recognition of small molecules, making the material ideally suited for applications such as separations, chemical sensing and catalysis^{2–5}. Until now, the molecular imprinting of bulk polymers^{2–5} and polymer^{6,7} and silica^{8,9} surfaces has been reported, but the extension of these methods to a wider range of materials remains problematic. For example, the formation of substrate-specific cavities within bulk silica, while conceptually straightforward¹⁰, has been difficult to accomplish experimentally^{11,12}. Here we describe the imprinting of bulk amorphous silicas with single aromatic rings carrying up to three 3-aminopropyltriethoxysilane side groups; this generates and occupies microporosity and attaches functional organic groups to the pore walls in a controlled fashion. The triethoxysilane part of the molecules' side groups is incorporated into the silica framework during sol-gel synthesis, and subsequent removal of the aromatic core creates a cavity with spatially organized aminopropyl groups covalently anchored to the pore walls. We find that the imprinted silicas act as shape-selective base catalysts. Our strategy can be extended to imprint other functional groups, which should give access to a wide range of functionalized materials.

One of the main difficulties in imprinting syntheses that use non-covalent interactions is heterogeneity in the imprinted sites that are formed^{2,13}. Because of the likelihood of site heterogeneity with non-covalent imprinting, we developed an imprinting methodology for bulk silica that involves covalent interactions. Figure 1 illustrates our general imprinting strategy, which we use here to synthesize bulk, amorphous, microporous silicas with binding sites comprising either one, two or three propylamine groups that are covalently linked to the silicon atoms of the silicate. In order to create binding sites that contain aminopropyl groups, the amine in 3-aminopropyltriethoxysilane is protected by converting it to a carbamate, as free amines can catalyse the sol-gel polymerization of tetraethoxysilane (TEOS) and precipitate the acid catalyst that is used for the hydrolysis of the ethoxysilanes. In addition to chemically protecting

the primary amine, the carbamate provides the linkage to the portion of the imprint that creates the void space (upon its removal) and provides spatial organization within the binding site (here it is a single aromatic ring). Imprint molecules containing one, two (in the para positions) and three (in the 1,3,5 positions as shown in Fig. 1) protected 3-aminopropyltriethoxysilane groups were prepared (see Supplementary Information for schematic representations and NMR and chemical analysis data). The imprint molecules were then reacted with TEOS in an acid-catalysed sol-gel copolymerization, as schematically illustrated in Fig. 1, to generate what we denote as one-, two- and three-point imprinted materials, respectively. These amorphous (verified by X-ray powder diffraction) and optically transparent organic-inorganic, hybrid materials were reacted with an equimolar mixture of chlorotrimethylsilane and 1,1,1,3,3,3-hexamethyldisilazane to cap surface silanol groups. Next, the C–N bonds of the imprint were cleaved by reaction with trimethylsilyliodide, and the resulting trimethylsilylcarbamate was exposed to methanol and water to liberate the imprinted propylamine sites.

Figure 2 shows the ¹³C cross polarization (CP) NMR and ²⁹Si CP NMR spectra from the two-point material (see Supplementary Information for spectra of the one- and three-point materials). The organic groups are covalently linked to the silicate silicon atoms

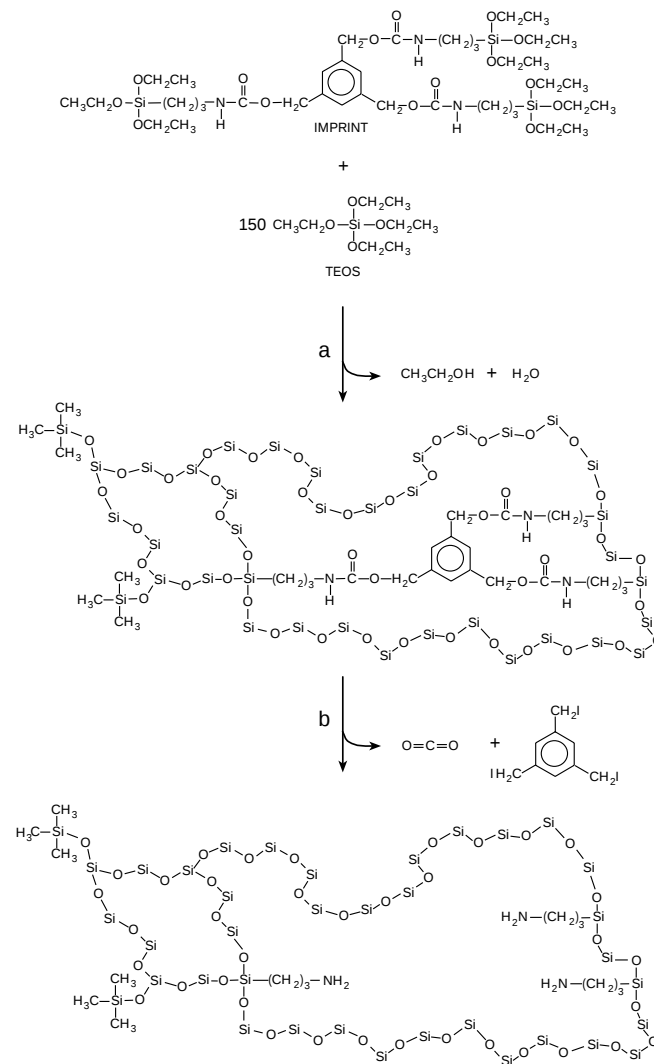


Figure 1 Preparation procedures used to create the imprinted silicas. **a**, sol-gel hydrolysis and condensation catalysed by HCl (pH 2.0). **b**, First, treatment with trimethylsilyliodide in acetonitrile (0.25 M) and second, wash with methanol/aqueous sodium bicarbonate.

[†] Present address: Chemical Engineering Department, University of California, Berkeley, California 94720-1462, USA.

(peak at -68 ppm marked with an asterisk indicates $\text{C-Si}(\text{OSi})_3$, ref. 14 and references therein; and $(\text{CH}_3)_3\text{Si-O-Si}$ gives chemical shifts that are positive). Also, the local atomic arrangement of silicon atoms does not change upon imprint removal (the spectra in Fig. 2a and b are the same). The ^{13}C CP NMR spectra in Fig. 2c and d show that the imprint is completely intact in the as-synthesized material and that most of the C-N bonds can be cleaved during the imprint removal process. Although it is not possible to assign a quantitative value from a ^{13}C CP NMR spectrum, it is estimated that $85 \pm 10\%$ of the imprint molecules are removed from the two-point material, thus giving 0.27 mmol of amine per gram. The primary amines of the organic-functionalized silicas are able to form Schiff bases with acetylacetone (observed by diffuse reflectance ultraviolet band at 316 nm), and the site densities determined by adsorption of benzoic acid are 0.23 , 0.23 and 0.07 mmol amine per gram corresponding to 73% , 73% and 22% extent of imprint removal for the one-, two- and three-point materials, respectively (it is significantly more difficult to remove imprint from the three-point material). A control silica prepared by the synthesis procedure used for the imprinted materials (no imprint used) gave a value of 0.02 mmol benzoic acid per gram for comparison, presumably due to acid that remains physically occluded in the micropores. Figure 3 shows the results from the physical adsorption of argon at 77 K for the two-point material. Isotherm (I) of Fig. 3a shows that the as-synthesized solid contains porosity and that the majority of the pore volume is in the micropore region (Fig. 3b). Acid-catalysed sol-gel reactions of

TEOS are known to create micropores that are filled with solvent (mainly ethanol here)¹⁵. Upon removal of the imprint, additional porosity is created (Fig. 3a) that is confined to the micropores (the difference in isotherms (I) and (II) is below $P/P_0 \approx 0.1$, see inset to Fig. 3b, and the pore volume created is approximately 0.8 – 1.0 nm in size, Fig. 3b) and is of the expected size based on the imprint molecule used. Here P is the pressure of N_2 and P_0 is the saturation pressure for N_2 , which is 1 atm. We estimated the volume of the organic fragment that is removed from molecular modelling, and hence the physical adsorption data reveal that $\sim 80\%$ of the imprint is lost and is consistent with the values obtained by NMR and chemisorption. This is the first time, to our knowledge, that molecular-imprint-generated microporosity has been directly observed using a physical adsorption experiment. The two-point material is an active base catalyst for the Knoevenagel condensation reaction of malononitrile with isophthalaldehyde (74 turnovers per amine site per hour). The two-point material is shape-selective in that the addition of the second malononitrile is inhibited while it freely occurs using aminopropyl groups on the surface of amorphous silica (367 turnovers per amine site per hour). These results are evidence that the propylamine groups in the imprinted silicas

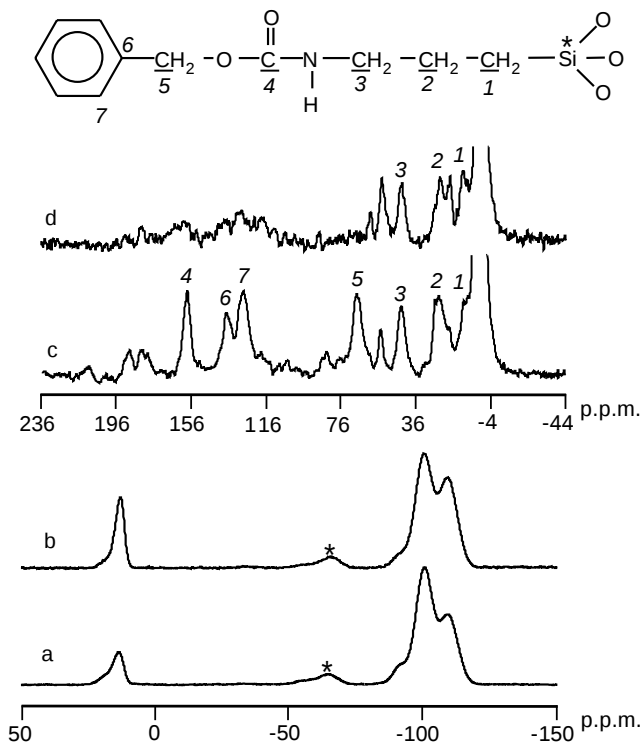


Figure 2 Solid-state NMR spectra of the two-point-imprinted material. **a**, ^{29}Si cross polarization (CP) NMR spectrum (as-synthesized and reacted with a mixture of chlorotrimethylsilane and 1,1,1,3,3,3-hexamethyldisilazane). **b**, ^{29}Si CP NMR spectrum of material after imprint removal. **c**, ^{13}C CP NMR spectrum of solid in **a**. **d**, ^{13}C CP NMR spectrum of material in **b**. Solid-state NMR spectroscopy was performed on a Bruker AM 300 spectrometer equipped with solids accessories. Samples were packed into 7-mm ZrO_2 rotors and spun in air at approximately 4 kHz. Typical recycle delays were 60 seconds. The resonances between peaks 1 and 2 and between 3 and 5 in **c** and **d** are from a small amount of residual ethoxy groups.

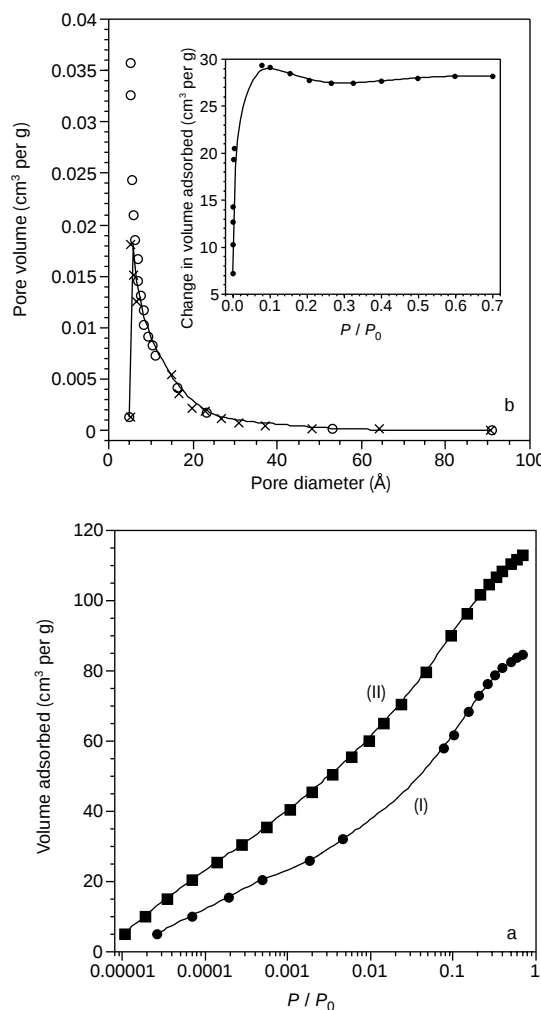


Figure 3 Physical adsorption of argon at 77 K. **a**, Adsorption isotherms for (I) the as-synthesized two-point material and (II) the material used in (I) after imprint removal. **b**, Pore volume distribution from isotherm (I) (crosses) and isotherm (II) (circles). Inset, difference isotherm between isotherms (I) and (II). Measurements were performed using a Micromeritics ASAP 2010 system on samples that had been vacuum degassed at a temperature of 363 K.

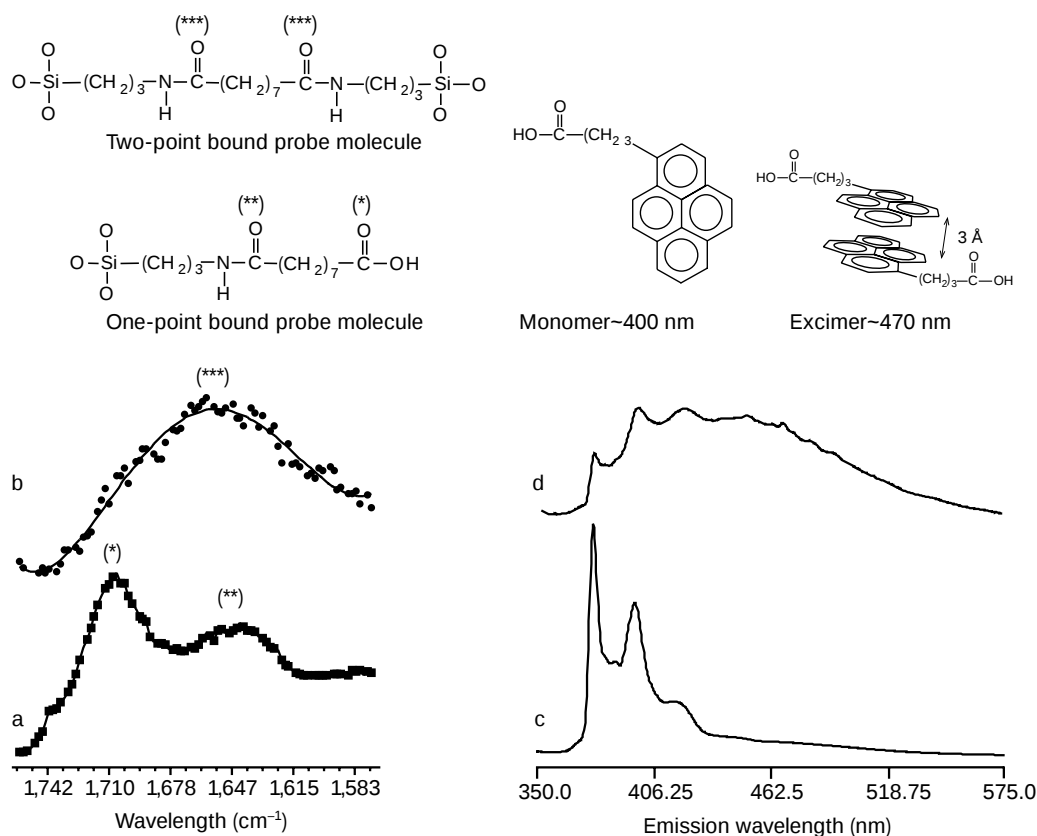


Figure 4 Infrared and fluorescence spectra from imprinted silicas. **a, b**, Fourier-transform infrared spectra of one-point (**a**) and two-point (**b**) materials contacted with azelaoyl chloride. **c, d**, Fluorescence emission spectra of one-point (**c**) and three-point (**d**) materials contacted with pyrenebutyric acid.

are residing in pores of the silica and that they can accomplish a liquid-phase, shape-selective, base-catalysed reaction.

It is clear from the data provided above that organic-functionalized, amorphous, microporous silicas can be prepared using this method. To address the spatial positioning of the amine groups, two different experiments were performed. The one- and two-point materials were exposed to azelaoyl chloride in cyclohexane or chloroform at room temperature and the resulting solids washed to remove physisorbed material. This molecule is the correct length to form a diamide in the binding site of the two-point material, and only a single amide in the one-point material (thus exhibiting site isolation)¹⁶. Approximately 35% of the binding sites were adsorbed with the probe molecule in both silicas (see Supplementary Information for procedure). The Fourier-transform infrared spectra shown in Fig. 4 reveal the selective binding capability of the imprinted materials. Amide formation on the two-point material (amide I band at 1,630–1,665 cm^{-1}) is observed and the one-point material shows both an amide (amide I band at 1,630–1,655 cm^{-1}) and a carboxylic acid group (at 1,700 cm^{-1} ; the band shifts to 1,575 cm^{-1} upon exposure to triethylamine, indicative of the acid–amine interaction to form a carboxylate salt). The absorption from the two-point material is broad and may be due to the rigidity of the diamide species, as amide vibrations from the imprints in the as-synthesized silicas are also broad. Pyrenebutyric acid (PBA) was contacted with the one- and three-point materials, after which the solids were Soxhlet extracted with chloroform for 24 hours to remove physisorbed PBA (see Supplementary Information for procedure). The fluorescence emission spectra from these materials are shown in Fig. 4. The one-point material reveals emission mainly from a monomer whereas the three-point material gives significant excimer formation. The control silica (no amine) was contacted with PBA and the emission spectrum obtained in order to observe

the behaviour from occluded PBA (ref. 17). In addition, amorphous silica with surface aminopropyl groups (same material as used for catalysis) was investigated to show the effects of surface amines. Monomer and excimer emission are observed on all samples. The ratio of monomer to excimer in the one- and three-point materials is 17 and 0.7, respectively, while for the control silica the ratio is 1.3 (physically adsorbed PBA) and for the surface-bound amine silica it is 6.7. Thus, site isolation (one-point) and clustering (three-point) are observed when expected and are each significantly different from the appropriate controls.

Methods

Synthesis of 3-(triethoxysilyl)propyl-benzylcarbamate

To 42 mmol (10 ml) of 3-aminopropyltriethoxysilane in 50 ml of anhydrous ether, 40 mmol (5.6 ml) of triethylamine were added. Subsequently, 40 mmol (5.7 ml) of benzyl chloroformate were added dropwise with a dry syringe. The mixture was stirred at room temperature for 1 hour, redissolved in chloroform, and extracted with pH 2.0 aqueous HCl, saturated sodium bicarbonate, and saturated brine. The organic phase was purified by flash chromatography to a transparent oil (yield 80%).

Synthesis of 3-(triethoxysilyl)propyl-1,4-phenylenebis(methylene)carbamate

To 24 mmol (3.89 g) of 1,1'-carbonyldiimidazole in 20 ml of dry 1,2,3,4-tetrahydro-9-fluorenone (THF), a catalytic amount of sodium ethoxide was added. Subsequently 12 mmol (1.66 g) of 1,4-benzenedimethanol were dissolved in 20 ml of dry THF and were added to the reaction mixture, followed by the addition of 26 mmol (6.2 ml) of 3-aminopropyltriethoxysilane. The mixture was stirred at room temperature and was monitored by thin-layer chromatography. The crude product was redissolved in ether, and excess imidazole was removed by filtration. Purification by silica chromatography (Silica Gel 60 and 1.4/1.0 v/v hexane:ethyl acetate) and recrystallization in chloroform/hexane resulted in white platelets (yield 41%).

Synthesis of [3-(triethoxysilyl)propyl]-1,3,5-benzenetriyltris(methylene)carbamate

To 36 mmol (5.84 g) of 1,1'-carbonyldiimidazole in 20 ml of dry THF, a catalytic amount

of sodium ethoxide was added. Subsequently 12 mmol (2.02 g) of 1,3,5-benzenetri-methanol (synthesized as described previously¹⁸) were dissolved in 20 ml of dry THF. The reaction mixture was stirred at room temperature for approximately 2 hours. Subsequently 40 mmol (9.3 ml) of 3-aminopropyltriethoxysilane were added. The reaction mixture was allowed to stir at room temperature and was monitored by thin-layer chromatography. The crude product was redissolved in ether, and excess imidazole was removed via filtration. Purification by silica chromatography (Silica Gel 60 and 1.4/1.0 v/v hexane:ethyl acetate) resulted in a white wax-like solid (yield 55%).

Synthesis of silica

The amount of imprint used in the sol-gel synthesis corresponds to 2 mol% of imprint silicon relative to TEOS silicon. A typical procedure was to dissolve 0.7338 g of one-point imprint for one-point material synthesis (no imprint for control, 0.6532 g of two-point imprint and 0.6263 g of three-point imprint) into 193.5 ml of dry ethanol in a 16-ounce jar (Qorpak 7534). 23.0 ml of TEOS were added to this mixture. Finally 64.5 ml of pH 2.0 aqueous HCl were added to the gel mixture. The mixture was covered with a loose cap, stirred for 24 hours at 8 °C, stirred for 12 hours at 15 °C and for 8 days at room temperature. This mixture was transferred to a 40 °C oven. The mixture was aged in the oven for 10 days after which time gelation had occurred. The resulting glass monoliths were aged for 8 days at 40 °C. The final gels had an approximate imprint content of 0.283 mmol imprint nitrogen per gram of as-made material as determined from the mass balance. A thermogravimetric analysis (TGA) showed that the gels can be heated to over 200 °C in air without any detectable decomposition of imprint. The as-made imprinted silica monoliths were ground into a powder of particles less than 10 µm in diameter (verified by scanning electron microscopy). The powder was Soxhlet extracted with acetonitrile refluxing in calcium hydride. The silica was then separately washed with chloroform and pentane and allowed to dry. The silica was subsequently treated with an equimolar mixture of neat chlorotrimethylsilane and 1,1,1,3,3,3-hexamethyldisilazane at room temperature for 24 hours and finally was washed with dry THF, dry acetonitrile, chloroform, and pentane.

Procedures for imprint cleavage

15 ml per gram of silica of a 0.25 M trimethylsilyliodide in dry acetonitrile solution were added under argon to a batch of capped silica, and the resulting slurry was stirred in the dark for 12 hours under argon. Increasingly higher temperatures were used to effect deprotection in the multiple point materials; thus, 40 °C for the one-point material, 70 °C for the two-point material and 80 °C for the three-point material were used. The deprotected silica was collected by vacuum filtration and washed with dry acetonitrile, methanol, saturated aqueous sodium bicarbonate, methanol and acetonitrile. The deprotected silica was then Soxhlet extracted with acetonitrile refluxing in calcium hydride and chloroform and finally washed with pentane.

Catalytic reactions

1.5 mmol isophthalaldehyde and 3.0 mmol malononitrile were combined in 5ml acetonitrile containing 8 mg hexamethylbenzene as an internal standard and heated to 80 °C. Subsequently, 30 mg of imprinted silica (0.2 mmol amine per gram) or 6 mg of a surface-functionalized, amorphous silica (1 mmol amine per gram) were added and the progress of the reaction was monitored by gas chromatographic analysis.

Received 22 June; accepted 1 December 1999.

- Whitcombe, M. J., Rodriguez, M. E., Villar, P. & Vulfson, E. A new method for the introduction of recognition site functionality into polymers prepared by molecular imprinting - Synthesis and characterization of polymer receptors for cholesterol. *J. Am. Chem. Soc.* **117**, 7105-7111 (1995).
- Vlatakis, G., Anderson, L. I., Müller, R. & Mosbach, K. Drug assay using antibody mimics made by molecular imprinting. *Nature* **361**, 645-647 (1993).
- Wulff, G. Molecular imprinting in cross-linked materials with the aid of molecular templates - a way towards artificial antibodies. *Angew. Chem. Int. Edn Engl.* **34**, 1812-1832 (1995).
- Davis, M. E., Katz, A. & Ahmad, W. R. Rational catalyst design via imprinted nanostructured materials. *Chem. Mater.* **8**, 1820-1839 (1996).
- Shea, K. J. Molecular imprinting of synthetic network polymers: the de novo synthesis of macro-molecular binding and catalytic sites. *Trends Polym. Sci.* **2**, 166-173 (1994).
- D'Souza, S. M. *et al.* Directed nucleation of calcite at a crystal-imprinted polymer surface. *Nature* **398**, 312-316 (1999).
- Shi, H., Tsai, W., Garrison, M. D., Ferrari, S. & Ratner, B. D. Template-imprinted nanostructured surfaces for protein recognition. *Nature* **398**, 593-597 (1999).
- Wulff, G., Heide, B. & Helfmeier, G. Molecular recognition through the exact placement of functional groups on rigid matrices via a template approach. *J. Am. Chem. Soc.* **108**, 1089-1091 (1986).
- Dai, S. *et al.* Imprint coating: A novel synthesis of selective functionalized ordered mesoporous sorbents. *Angew. Chem. Int. Edn Engl.* **38**, 1235-1239 (1999).
- Heilmann, J. & Maier, W. F. Selective catalysis on silicon dioxide with substrate-specific cavities. *Angew. Chem. Int. Edn Engl.* **33**, 471-473 (1994).
- Ahmad, W. R. & Davis, M. E. Transesterification on "imprinted" silica. *Catal. Lett.* **40**, 109-114 (1996).
- Maier, W. F. & Ben Mustapha, W. Transesterification on "imprinted" silica—Reply. *Catal. Lett.* **46**, 137-140 (1997).
- Katz, A. & Davis, M. E. Investigations into the mechanisms of molecular recognition with imprinted polymers. *Macromolecules* **32**, 4113-4121 (1999).
- Jones, C. W., Tsuiji, K. & Davis, M. E. Organic-functionalized molecular sieves as shape-selective catalysts. *Nature* **393**, 52-54 (1998).
- Brinker, C. J., Keefer, K. D., Schaefer, D. W. & Ashley, C. S. Sol-gel transition in simple silicates. *J. Non-Cryst. Solids* **48**, 47-64 (1982).

- Fuji, K. *et al.* Visualization of molecular length of α,ω -diamines and temperature by a receptor based on phenolphthalein and crown ether. *J. Am. Chem. Soc.* **121**, 3807-3808 (1999).
- Dunn, D. & Zink, J. I. Probes of pore environment and molecule-matrix interaction in sol-gel materials. *Chem. Mater.* **4**, 2280-2291 (1997).
- Houk, J. & Whitesides, G. M. Structure reactivity relations for thiol disulfide interchange. *J. Am. Chem. Soc.* **109**, 6825-6836 (1987).

Acknowledgements

A Fannie and John Hertz Foundation Graduate Fellowship to A. K. and a NSF Waterman Award to M. E. D. are gratefully acknowledged. We thank L. W. Beck for assistance with the NMR experiments, and H. Gonzalez and A. G. Myers for discussions.

Correspondence and requests for materials should be addressed to M. E. D. (e-mail: mdavis@chemo.caltech.edu).

Biomimetic synthesis of ordered silica structures mediated by block copolypeptides

Jennifer N. Cha*, Galen D. Stucky†*, Daniel E. Morse‡ & Timothy J. Deming†*

Departments of * Chemistry, † Materials, and ‡ Molecular, Cellular and Developmental Biology, University of California, Santa Barbara, California 93106, USA

In biological systems such as diatoms and sponges, the formation of solid silica structures with precisely controlled morphologies is directed by proteins and polysaccharides and occurs in water at neutral pH and ambient temperature¹⁻⁴. Laboratory methods, in contrast, have to rely on extreme pH conditions and/or surfactants to induce the condensation of silica precursors into specific morphologies or patterned structures⁵⁻¹⁰. This contrast in processing conditions and the growing demand for benign synthesis methods that minimize adverse environmental effects have spurred much interest in biomimetic approaches in materials science^{4,5}. The recent demonstration that silicatein—a protein found in the silica spicules of the sponge *Tethya aurantia*¹¹—can hydrolyse and condense the precursor molecule tetraethoxysilane to form silica structures with controlled shapes at ambient conditions¹²⁻¹⁴ seems particularly promising in this context. Here we describe synthetic cysteine-lysine block copolypeptides that mimic the properties of silicatein: the copolypeptides self-assemble into structured aggregates that hydrolyse tetraethoxysilane while simultaneously directing the formation of ordered silica morphologies. We find that oxidation of the cysteine sulphhydryl groups, which is known to affect the assembly of the block copolypeptide¹⁵, allows us to produce different structures: hard silica spheres and well-defined columns of amorphous silica are produced using the fully reduced and the oxidized forms of the copolymer, respectively.

Because tetraethoxysilane (TEOS) is stable when mixed with water at neutral pH, a successful biomimetic silica synthesis from this precursor requires an agent that displays hydrolytic activity simultaneously with structure-directing properties. Site-directed mutagenesis of the cloned DNA coding for silicatein- α revealed that interacting histidine and serine residues were required for the hydrolytic activity of this protein^{4,14}. Based on this precedent, we evaluated simple homopolypeptides of amino acids bearing polar functional groups for their ability to mimic the properties of silicatein in the polycondensation of silicon alkoxides. When the homopolymers of L-lysine, L-histidine, D/L-serine, L-threonine and L-glutamic acid were separately dissolved in aqueous pH 7 buffer

Table 1 Block copolypeptides screened for their ability to react with TEOS to form silica

Block copolymer	Composition	Synthesis yield (%)	SiO ₂ rate		Shape	
			N ₂	Air	N ₂	Air
1	poly(L-alanine ₃₀ -b-L-lysine ₂₀₀)	87	0.09(2)	NA	N	NA
2	poly(L-glutamine ₃₀ -b-L-lysine ₂₀₀)	88	0.22(7)	NA	N	NA
3	poly(L-serine ₃₀ -b-L-lysine ₂₀₀)	84	0.29(3)	NA	N	NA
4	poly(L-tyrosine ₃₀ -b-L-lysine ₂₀₀)	86	0.30(7)	NA	N	NA
5	poly(L-cysteine ₁₀ -b-L-lysine ₂₀₀)	76	0.60(2)	0.60(2)	S	S
6	poly(L-cysteine ₃₀ -b-L-lysine ₂₀₀)	77	0.43(2)	0.62(4)	S	C
7	poly(L-cysteine ₆₀ -b-L-lysine ₂₀₀)	87	0.37(4)	0.67(1)	E	C
8	poly(L-cysteine ₃₀ -b-L-lysine ₄₀₀)	88	0.62(1)	0.65(4)	S	S
9	poly(L-cysteine ₃₀ -b-L-glutamate ₂₀₀)	90	0.01(1)	0.01(1)	NA	NA
10	poly(L-cysteine ₃₀)	76	0.43(6)	0.08(1)	N	N
11	poly(L-lysine ₂₀₀)	96	0.01(1)	NA	NA	NA
12	none	NA	0.01(1)	NA	NA	NA

Synthesis yield, total isolated yield of deprotected copolymer; SiO₂ rate, initial rate of silica formation (in $\mu\text{mol h}^{-1}$) mediated by block copolypeptide at a concentration of 5 mg ml⁻¹ in 50 mM Tris-HCl buffer, pH 6.8 and an initial TEOS concentration of 3.4 M. The silica precipitate was collected by centrifugation, washed with 95% ethanol and dissolved in 0.2 M NaOH at 37 °C. The amount of silica was then determined using the spectrophotometric molybdate assay^{24,25}. N₂ indicates that silica preparation was carried out under an oxygen-free nitrogen atmosphere; Air indicates that silica preparation was carried out in air. In the absence of polymer, buffer alone was used as the control. Shape, morphology of silica particles: N, non-ordered; S, spheres; E, elongated globules; C, columns. NA, not applicable; b, block copolymer.

and mixed with TEOS, it was found that none of these polymers was able to produce silica at ambient temperature over a 24-hour period. Furthermore, mixtures of these homopolymers also failed to catalyse TEOS hydrolysis and condensation. In contrast, we found that oligomers of L-cysteine (of molecular mass 3000 Da, used because higher chain lengths were insoluble) efficiently produce silica from TEOS in pH 7 buffer (Table 1), when handled under an inert nitrogen atmosphere to prevent oxidation. This result was presumably due to the nucleophilic properties of the sulphhydryl group, which may enable it to initiate hydrolysis of the silicon alkoxide. When these oligomers were used under air, oxidation of the sulphhydryl groups to disulphides resulted in insoluble aggregates that were much less active in silica formation (Table 1)¹⁶. However, the silica formed by using oligo-L-cysteine was an amorphous powder with no defined macroscopic shape. From these results we concluded that simple homopolymers of amino acids, which lack the structural complexity and polyfunctionality found in proteins, are unable to reproduce the shape-controlling ability of silicatein^{4,11–14}.

In an effort to better mimic this protein, we synthesized diblock copolypeptides that contained covalently linked domains (blocks) of water-soluble and water-insoluble polypeptides. Dissimilarity in the block segments gave the chains an amphiphilic character, similar to that of surfactants, which resulted in self-assembly of the chains in aqueous solution¹⁷. The architecture and design of the block copolypeptides also provided a means to solubilize water-insoluble domains; for example, hydrolytically active poly-L-cysteine. For

these reasons, block copolypeptides were expected to allow the directed cooperative assembly, hydrolysis and condensation of TEOS to form specific silica structures. The solubilizing block copolypeptide components were either cationic or anionic polyelectrolytes, such as poly-L-lysine and poly-L-glutamate, which are known to be water soluble at pH 7 (ref. 18). As water-insoluble domains, poly-L-cysteine and poly-L-serine were chosen both for their potential silica-forming hydrolytic activity as well as their ability to aggregate in water (by either hydrogen or covalent bonding via β -sheet formation or disulphide linkages¹⁶). Other insoluble domains chosen included polar residues that were less nucleophilic than cysteine (poly-L-glutamine and poly-L-tyrosine) or slightly hydrophobic (poly-L-alanine). Poly-L-histidine was not used in these studies because of difficulty in protecting the side-chain to form a suitable amino acid N-carboxyanhydride (amino acid NCA) monomer.

The block copolypeptides that were synthesized and studied are given in Table 1. They were prepared from suitably protected amino acid NCA monomers by using the initiator 2,2'-bipyridylNi(1,5-cyclooctadiene). This synthetic protocol has been shown to give block copolypeptides of narrow molecular mass distributions and with controlled molecular masses¹⁹ (Fig. 1).

The cationic block copolymers showed more activity in silica formation than the corresponding anionic copolymer. In fact, poly-L-glutamate completely inhibited the ability of the poly-cysteine block to form silica, which supports the hypothesis that polycations are important for interacting with negatively charged silicate

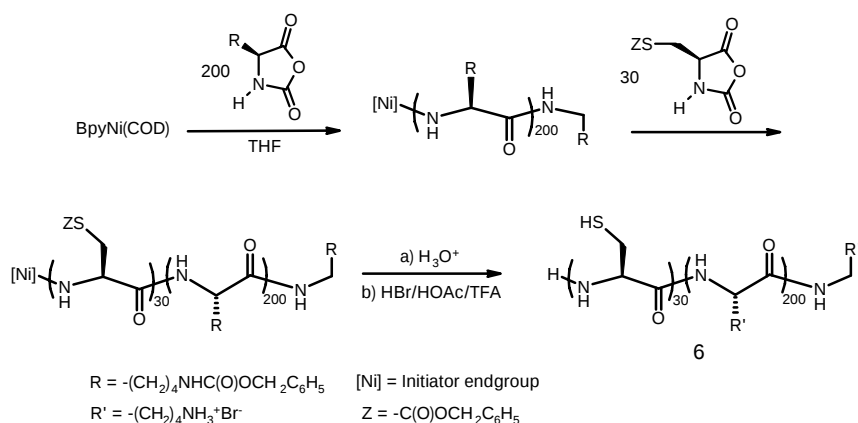


Figure 1 Synthesis strategy to produce block copolypeptides. Stepwise polymerization of the monomer, N_ε-carboxybenzyl-L-lysine NCA, followed by S-carboxybenzyl-L-cysteine NCA gave the protected polymer that was then deprotected using equimolar amounts of trifluoroacetic acid and 33% HBr in acetic acid to give **6**. BpyNi(COD) indicates 2,2'-

bipyridylnickel(1,5-cyclooctadiene). The protected copolymer was analysed using size-exclusion chromatography in dimethylformamide (DMF) at 60 °C to verify the molecular mass. Polymer composition was verified by ¹H NMR analysis of the deprotected copolymer in trifluoroacetic acid (TFA)-d

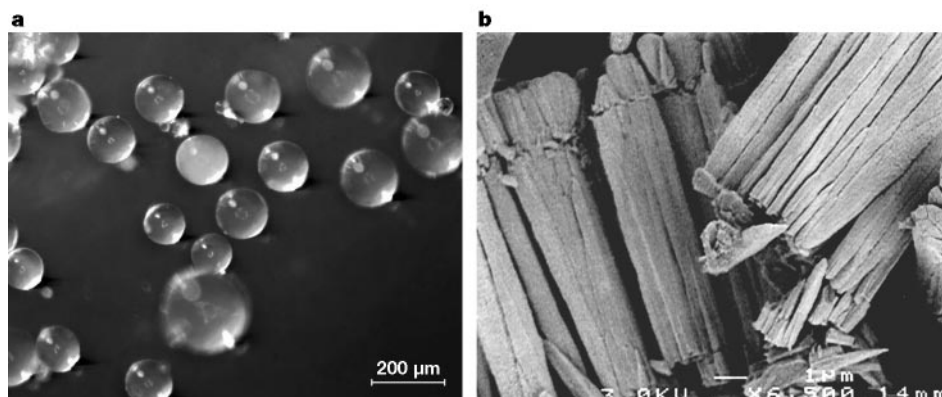


Figure 2 Different ordered silica shapes obtained using block copolypeptide **6**. In a typical procedure, TEOS (2.0 ml) was added to 500 μl of a solution of **6** (5 mg ml⁻¹ in 50 mM Tris-HCl buffer, pH 6.8), and the resulting biphasic mixture was agitated vigorously and then allowed to stand for several hours with no stirring, whereupon some of the TEOS had emulsified into the aqueous phase. After 24 h, the resulting silica precipitate was collected from the aqueous phase, washed with 95% ethanol and air dried. **a**, Optical micrograph of

silica spheres obtained from synthesis under nitrogen; scale bar, 200 μm. **b**, Scanning electron micrograph of packed silica columns obtained from synthesis under air; scale bar, 1 μm. The sample was sputter coated with gold and examined with a JEOL JSM 6300F equipped with a cold cathode field-emission source operated at a beam energy of 3.0 kV.

precursors²⁰. All of the lysine-containing copolymers displayed some activity in silica formation, and the rate of silica production increased steadily as the domain bound to the poly-L-lysine block became more nucleophilic. As polymer **1** (see Table 1), which contains no nucleophilic component, was able to produce silica, it appeared that poly-L-lysine itself, when constrained in a self-assembling block copolymer, possessed a low activity toward the hydrolysis and condensation of TEOS. However, the cysteine- and serine-containing copolymers were the only ones that were able to control the shape of the silica during its formation, with the cysteine-containing polymers being most active. For these reasons, further studies were focused on the cysteine-lysine block copolymer combination.

In initial experiments, polymer **6** (polymers **1**–**12** are defined in Table 1) was deprotected and handled under a nitrogen atmosphere, and thus was used in its reduced form when reacted with TEOS. Dynamic light scattering measurements of **6** in aqueous solution (1.6 mg ml⁻¹) showed that this polymer self-assembled into large aggregates approximately 600 nm in diameter. This colloidal polymer solution, when mixed with TEOS, formed a two-phase system in which some TEOS was emulsified into the aqueous phase. After 24 hours, the formation of transparent, composite silica spheres (diameter, ~ 100 μm) was observed (Fig. 2). ²⁹Si magic-angle-spinning NMR measurements confirmed the existence of highly condensed silica (35% Q³ and 65% Q⁴ species). When calcined at 500 °C, the spheres remained both intact and transparent without a decrease in apparent size, although thermal gravimetric analysis revealed a 10% weight loss of organic material. Nitrogen sorption measurements using the BET (Brunauer–Emmett–Teller) showed that the spheres were mesoporous, with a broad distribution of pore sizes and a surface area of 436 m² g⁻¹ (ref. 21). Using the block copolypeptide **6** to prepare these hard, transparent, mesoporous silica spheres represents (to our knowledge) the first example in which hydrolysis and condensation of an inorganic phase as well as structural templating were all controlled by a single synthetic material at pH 7, thus mimicking biological silica synthesis. As far as we know, other surfactant or polymer based systems developed for shape-selective silica synthesis typically require use of a catalyst and extreme pH conditions^{10,22,23}.

Copolymers similar to **6**, but with different block lengths, were also synthesized to determine the role of copolymer composition on silica-forming ability. A polymer with a shorter cysteine domain, **7**, was similar to **6** in being able to produce silica spheres. However, when the length of the cysteine domain was increased (**8**), the

formation of more elongated silica particles was observed. Increasing the size of the lysine domain gave a polymer, **9**, which behaved in the same way as the smaller, but similar-composition, **7**. This indicates that the copolymer chain-length has little effect on resulting silica shape. It should be noted that mixtures of L-lysine and L-cysteine homopolymers, in proportions similar to those found in the block copolypeptides **6**–**9**, only gave completely disordered silica powders from TEOS.

An additional feature of the cysteine residues in **6** was their ability to form covalent disulphide bonds as inter- and intra-chain crosslinks upon oxidation of the sulphhydryl groups. After deprotection of the copolymer in air, the formation of such disulphide crosslinks in oxidized **6** was evident from the high viscosity exhibited by this sample upon exposure to water¹⁵. The gel dissolved readily upon addition of a reducing agent such as β-mercaptoethanol, indicating the presence of disulphide crosslinks¹⁵. Dynamic light scattering measurements of oxidized solutions of **6** showed that the block copolymer aggregates had increased in size (diameter ~ 1,300 nm) relative to unoxidized samples (~ 600 nm). When oxidized **6** was mixed with TEOS, the rate of silica formation was found to increase, although 70% of the sulphhydryl groups had been converted to disulphide linkages (Table 1). In addition, ordered columns of silica were observed instead of spheres, showing that oxidation of the poly-L-cysteine domains was sufficient to completely modify the resulting topology of the silica (Fig. 2). With copolymers of different composition (Table 1), it could be shown that a minimum fraction of cysteine (~15 mol%) was required to produce the columnar silica composites. These results illustrate the importance of the self-assembled block copolypeptide architecture in the formation of silica shapes. The synthetic capability to control directly silica shape, hydrolysis and condensation rate by adjustment of block copolypeptide composition that we report here offers a route to the environmentally benign, biomimetic synthesis of inorganic materials. □

Received 23 July; accepted 3 December 1999.

1. Simpson, T. L. & Volcani, B. E. *Silicon and Siliceous Structures in Biological Systems* (Springer, New York, 1981).
2. Kröger, N., Lehmann, G., Rachel, R. & Sumper, M. Characterization of a 200-kDa diatom protein that is specifically associated with a silica-based substructure of the cell wall. *Eur. J. Biochem.* **250**, 99–105 (1997).
3. Vrieling, E. G., Beelen, T. P. M., van Santen, R. A. & Gieskes, W. W. C. Diatom silicon mineralization as an inspirational source of new approaches to silica production. *J. Biotechnol.* **70**, 39–51 (1999).
4. Morse, D. E. Silicon biotechnology: harnessing biological silica production to make new materials. *Trends Biotechnol.* **17**, 230–232 (1999).

- Mann, S. J. Biomimetic materials chemistry. *J. Mater. Chem.* **5**, 935–946 (1995).
- Oliver, S., Kuperman, A., Coombs, N., Lough, A. & Ozin, G. Lamellar aluminophosphates with surface patterns that mimic diatom and radiolarian microstructures. *Nature* **378**, 47–50 (1995).
- Monnier, A. *et al.* Cooperative formation of inorganic-organic interfaces in the synthesis of silicate mesostructures. *Science* **261**, 1299–1303 (1993).
- Zhao, D. Y. *et al.* Triblock copolymer syntheses of mesoporous silica with periodic 50 to 300 angstrom pores. *Science* **279**, 548–552 (1998).
- Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism. *Nature* **359**, 710–712 (1992).
- Ying, J. Y., Mehnert, C. P. & Wong, M. S. Synthesis and applications of supramolecular-templated mesoporous materials. *Angew. Chem. Int. Edn Engl.* **38**, 56–77 (1999).
- Shimizu, K., Cha, J., Stucky, G. D. & Morse, D. E. Silicatein alpha: cathepsin I-like protein in sponge biosilica. *Proc. Natl Acad. Sci. USA* **95**, 6234–6238 (1998).
- Cha, J. N. *et al.* Silicatein filaments and subunits from a marine sponge direct the polymerization of silica and silicones in vitro. *Proc. Natl Acad. Sci. USA* **96**, 361–365 (1999).
- Zhou, Y., Shimizu, K., Cha, J. N., Stucky, G. D. & Morse, D. E. Efficient catalysis of polysiloxane synthesis by silicatein alpha requires specific hydroxy and imidazole functionalities. *Angew. Chem. Int. Edn Engl.* **38**, 780–782 (1999).
- Morse, D. E. Silicon biotechnology: proteins, genes and molecular mechanisms controlling biosilica nanofabrication offer new routes to polysiloxane synthesis. *Organosilicon Chemistry IV: from Molecules to Materials* (eds Auner, N. & Weis, J.) (Wiley-VCH, New York, in the press).
- Liff, M. I. & Zimmerman, M. N. NMR Study of crosslinking by oxidation of four-cysteine polypeptide models of the elastic network phase of wool fibre. *Polym. Int.* **47**, 375–385 (1998).
- Berger, A., Noguchi, J. & Katchalski, E. Poly-L-cysteine. *J. Am. Chem. Soc.* **78**, 4483–4488 (1956).
- Zhang, L., Yu, K. & Eisenberg, A. Ion-induced morphological changes in crew-cut aggregates of amphiphilic block copolymers. *Science* **272**, 1777–1779 (1996).
- Bamford, C. H., Elliot, A. & Hanby, W. E. *Synthetic Polypeptides* (Academic, New York, 1956).
- Deming, T. J. Facile synthesis of block copolypeptides of defined architecture. *Nature* **390**, 386–389 (1997).
- Mizutani, T., Nagase, H., Fujiwara, N. & Ogoshi, H. Silicic acid polymerization catalyzed by amines and polyamines. *Bull. Chem. Soc. Jpn* **71**, 2017–2022 (1998).
- Lukens, W. W. Jr, Schmidt-Winkel, P., Zhao, D., Feng, J. & Stucky, G. D. Evaluating pore sizes in mesoporous materials: a simplified standard adsorption method and a simplified Broekhoff-de Boer method. *Langmuir* **15**, 5403–5409 (1999).
- Huo, Q. S., Feng, J. L., Schuth, F. & Stucky, G. D. Preparation of hard mesoporous silica spheres. *Chem. Mater.* **9**, 14–15 (1997).
- Zhao, D., Yang, P. D., Huo, Q. S., Chmelka, B. F. & Stucky, G. D. Topological construction of mesoporous materials. *Curr. Opin. Solid State Mater. Sci.* **3**, 111–121 (1998).
- Strickland, J. D. H. & Parsons, T. R. *A Practical Handbook of Seawater Analysis* 2nd edn (Fisheries Research Board of Canada, Ottawa, 1972).
- Brzezinski, M. A. & Nelson, D. M. A solvent-extraction method for the colorimetric determination of nanomolar concentrations of silicic-acid in seawater. *Mar. Chem.* **19**, 139–151 (1986).

Acknowledgements

We thank M. Brzezinski and B. Chmelka for suggestions, and E. Skogen for the light microscopy pictures. This work was supported by the US Army Research Office Multi-disciplinary University Research Initiative, the US Office of Naval Research, the NOAA National Sea Grant College Program, the US Department of Commerce, the California Sea Grant College System, the MRSEC Program of the NSF to the UCSB Materials Research Laboratory, and the Dow Corning Corporation.

Correspondence and requests for materials should be addressed to T. J. D. (e-mail: tjeming@mrl.ucsb.edu).

Natural methyl bromide and methyl chloride emissions from coastal salt marshes

Robert C. Rhew, Benjamin R. Miller & Ray F. Weiss

Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093-0244, USA

Atmospheric methyl bromide (CH_3Br) and methyl chloride (CH_3Cl), compounds that are involved in stratospheric ozone depletion, originate from both natural and anthropogenic sources. Current estimates of CH_3Br and CH_3Cl emissions from oceanic sources, terrestrial plants and fungi, biomass burning and anthropogenic inputs do not balance their losses owing to oxidation by hydroxyl radicals, oceanic degradation, and consumption in soils, suggesting that additional natural terrestrial sources may

be important¹. Here we show that CH_3Br and CH_3Cl are released to the atmosphere from all vegetation zones of two coastal salt marshes. We see very large fluxes of CH_3Br and CH_3Cl per unit area: up to 42 and 570 $\mu\text{mol m}^{-2} \text{d}^{-1}$, respectively. The fluxes show large diurnal, seasonal and spatial variabilities, but there is a strong correlation between the fluxes of CH_3Br and those of CH_3Cl , with an average molar flux ratio of roughly 1:20. If our measurements are typical of salt marshes globally, they suggest that such ecosystems, even though they constitute less than 0.1% of the global surface area², may produce roughly 10% of the total fluxes of atmospheric CH_3Br and CH_3Cl .

Field studies were conducted between June 1998 and June 1999 in two southern California coastal salt marshes: the Mission Bay marsh (32° 47' N, 117° 13' W) and the San Dieguito lagoon (32° 58' N, 117° 15' W). The climate at these locations is Mediterranean, with wet winters and dry summers. During the March to September growing season, the soils are typically hypersaline; this is because tidal sea water provides most of the soil moisture and because evaporation usually exceeds precipitation³. Study plots (Table 1) were chosen within predominant vegetation communities along a vertical zonation typical of California salt marshes, from the upper marsh (which is inundated only at the highest tides) down to the tidal channels which are devoid of vascular plants. Gas fluxes were measured using a dark static flux chamber⁴ which covers a surface area of 1 m^2 and encloses a volume of 850 litres; the large chamber size reduces possible 'edge effects' and allows for whole plants to be enclosed. Samples of air were extracted from the chamber, and

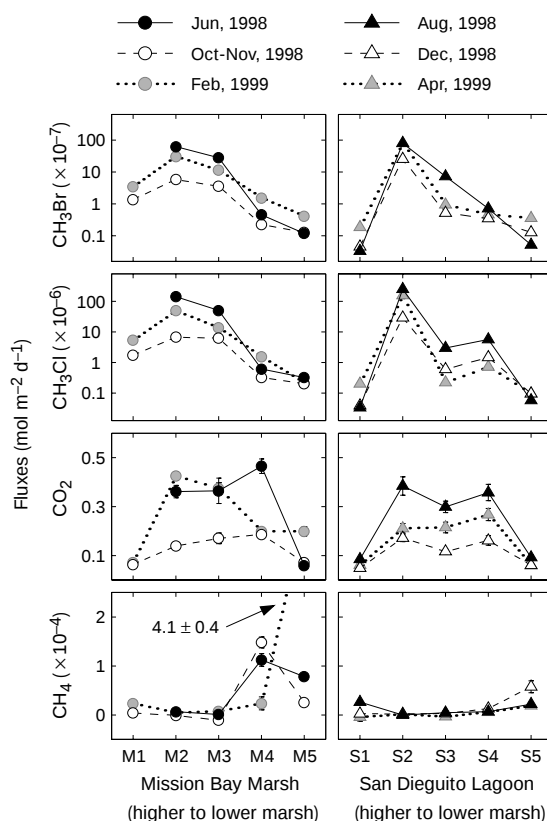


Figure 1 Daytime fluxes of CH_3Br , CH_3Cl , CO_2 and CH_4 . Measurements were made between June 1998 and April 1999 at the Mission Bay marsh (circles, left panels) and the San Dieguito lagoon (triangles, right panels). Fluxes are highly variable, both seasonally and spatially, with the largest CH_3Br and CH_3Cl emissions from the upper-middle marsh during the growing season (March–September). The predominant types of vegetation and flux data for each site are listed in Table 1. Note the logarithmic scales for the CH_3Br and CH_3Cl fluxes. The error bars for CO_2 and CH_4 fluxes include analytical and curve-fitting errors.

stored in stainless steel canisters that were lined with fused silica. The concentrations of CH₃Br, CH₃Cl, methane (CH₄) and carbon dioxide (CO₂) in these air samples were measured by gas chromatography. (See Methods for details.)

To quantify fluxes for each gas species, mole fractions within the chamber were plotted versus time of sampling, weighted linear least-squares fits were applied to these data, and the resulting slopes were multiplied by the number of moles of air in the chamber. Chamber measurements represent the net fluxes of the enclosed soil and plants, with positive slopes representing net production. Control experiments demonstrate no significant reactivity of the measured gases on the flux chamber surfaces over the length of the experiment, and the storage of air samples in the canisters has been tested for stability; there was no significant degradation, or production, of the compounds of interest⁵.

The results of our study of seasonal variations across the vegetation zones of these two marshes (Table 1, Fig. 1) show that CH₃Br and CH₃Cl are released from all salt-marsh sites, with greater emissions in the growing season than in the non-growing season. The magnitudes of these daytime fluxes, which vary widely from site to site, show patterns corresponding to the intertidal vegetation

zones of the salt marsh. The greatest emissions of the methyl halides are from the middle to upper-middle marsh, where *Salicornia* species, *Batis maritima*, and *Frankenia grandifolia* are the predominant vascular plants. Large methyl halide fluxes are typically associated with relatively high densities of above-ground green biomass, and do not seem to be inhibited by complete soil inundation. Fluxes at several sites decrease tenfold during the non-growing season, exhibiting a seasonality that corresponds with plant growth. These observations are consistent with the biogenic production of methyl halides by salt-marsh vegetation or microflora intimately associated with the plants. Also, sites with relatively large CH₄ fluxes (M4, M5, and S5; see Table 1) have low emissions of CH₃Br and CH₃Cl, suggesting that the highly reduced soils and/or vegetative cover of the lower marsh are less favourable for methyl halide emissions, even though the methyl halides are partially reduced compounds. In salt marshes, high CH₄ emissions are not only associated with lower-marsh *Spartina* grasses but are also typically associated with lower salinity^{6,7} which may affect methyl halide production rates.

The results from the upper-middle salt marsh (M2a, S2) during the growing season (April, June and August measurements)

Table 1 Measured CH₃Br and CH₃Cl fluxes by study site

Site	Predominant vegetation	Date (dd/mm/yy)	Hour*	T† (°C)	CH ₃ Br‡ (μmol m ⁻² d ⁻¹)	CH ₃ Cl‡ (μmol m ⁻² d ⁻¹)
Seasonal salt marsh flux study results, arranged from upper tidal zone to lower tidal zone						
Mission Bay marsh						
M1	<i>Monanthochloe littoralis</i> (shoregrass)	10/10/98	10	24	0.14 ± 0.03	1.7 ± 0.2
		23/02/99	16	26	0.35 ± 0.07	5.3 ± 0.6
M2a	<i>Batis maritima</i> / <i>Salicornia bigelovii</i> (saltwort/ annual pickleweed)	13/06/98	10	25	6.2 ± 2.1	140 ± 30
		10/10/98	08	19	0.58 ± 0.11	6.7 ± 1.0
M3	<i>Salicornia virginica</i> (common pickleweed)	23/02/99	13	31	3.09 ± 0.12	49 ± 2
		13/06/98	11	27	2.8 ± 0.7	51 ± 6
M4	<i>Spartina foliosa</i> (cordgrass)	10/10/98	09	21	0.36 ± 0.14	6.3 ± 1.6
		23/02/99	15	29	1.13 ± 0.07	13.9 ± 0.7
M5	Tidal creek/ <i>Enteromorpha</i> spp. (algal mat)	13/06/98	13	26	0.045 ± 0.003	0.62 ± 0.04
		10/10/98	06	16	0.0227 ± 0.0006	0.325 ± 0.012
		23/02/99¶	11	29	0.154 ± 0.004	1.56 ± 0.04
		13/06/98	14	29	0.0119 ± 0.0011	0.324 ± 0.017
		20/11/98	14	24	0.0130 ± 0.0006	0.203 ± 0.010
		23/02/99¶	11	29	0.041 ± 0.003	0.213 ± 0.011
San Dieguito lagoon						
S1	<i>Distichlis spicata</i> (saltgrass)	08/08/98	10	27	0.0034 ± 0.0009	0.034 ± 0.006
		05/12/98	15	17	0.0046 ± 0.0003	0.039 ± 0.002
S2	<i>Frankenia grandifolia</i> (alkali heath)	21/04/99¶	11	27	0.019 ± 0.003	0.20 ± 0.08
		08/08/98	06§	19	7.9 ± 0.7	250 ± 20
S3	<i>Salicornia subterminalis</i> (glasswort)	05/12/98	11§	18	2.5 ± 0.3	30 ± 3
		21/04/99	09	24	8.1 ± 2.7	160 ± 30
S4	<i>Salicornia virginica</i> (common pickleweed)	08/08/98	07§	19	0.73 ± 0.06	3.0 ± 0.2
		05/12/98	12§	17	0.052 ± 0.004	0.60 ± 0.05
S5	<i>Spartina foliosa</i> (cordgrass)	21/04/99	08	20	0.094 ± 0.008	0.22 ± 0.03
		08/08/98	08§	21	0.071 ± 0.007	5.7 ± 0.5
		05/12/98	13§	18	0.035 ± 0.004	1.5 ± 0.2
		21/04/99	10	28	0.047 ± 0.004	0.72 ± 0.07
		08/08/98	11§	29	0.0052 ± 0.0006	0.058 ± 0.005
		05/12/98	16§	14	0.0128 ± 0.0007	0.096 ± 0.006
		21/04/99¶	12§	23	0.036 ± 0.003	0.097 ± 0.018
		Diurnal salt marsh study results				
Mission Bay marsh						
M2a	See above	30/06/99	06	20	3.63 ± 0.14	78 ± 3
		30/06/99	13	27	10.0 ± 0.4	99 ± 4
		30/06/99	16	24	6.5 ± 0.2	65 ± 2
M2b	<i>S. bigelovii</i>	30/06/99	07	20	7.1 ± 0.3	155 ± 6
		30/06/99	12	26	13.5 ± 0.5	186 ± 7
		30/06/99	17	24	7.2 ± 0.3	103 ± 4
M2c	<i>B. maritima</i>	30/06/99	07§	21	11.8 ± 0.6	460 ± 60
		30/06/99	12§	26	42 ± 6	570 ± 30
		30/06/99	17§	23	29 ± 3	490 ± 20

* Pacific Standard Time (GMT minus 8 h).

† Average chamber air temperature.

‡ Flux errors (1σ) include analytical and curve-fitting errors.

§ Soils inundated.

¶ High *S. foliosa* mortality; *S. virginica* established at site.

¶ New site within 5 m of original site.

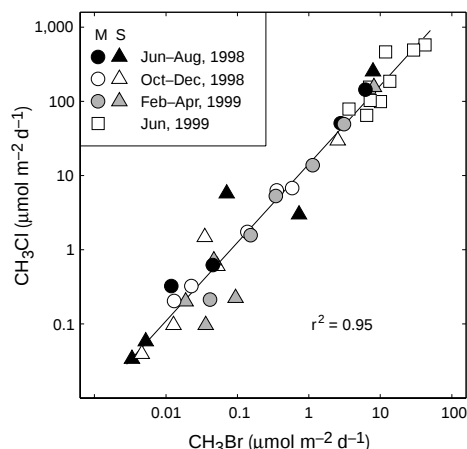


Figure 2 Correlation between CH_3Cl and CH_3Br fluxes. Fluxes are highly correlated over four orders of magnitude, despite the differences between Mission Bay marsh (M) and San Dieguito lagoon (S) sites, vegetation zones, and seasons of study. As both axes are on a logarithmic scale, the slope of near unity indicates that the fluxes are linearly correlated. The equation of the least-squares fitted line is: $\ln(\text{CH}_3\text{Cl flux}) = 1.06 \ln(\text{CH}_3\text{Br flux}) + 2.66$.

dominate the observed methyl halide fluxes. In order to characterize better the fluxes in this region, an additional study of growing-season diurnal and spatial variability was conducted at three upper-middle marsh sites at the Mission Bay marsh (Table 1). Experiments were 15 minutes in duration. These fluxes exhibit strong diurnal trends that follow the incident sunlight and air temperature changes, but not the soil surface (0–3 cm depth) temperature changes, suggesting that the methyl halides are emitted from the above-ground rather than subsurface biomass. The maximum observed fluxes per unit area of CH_3Br and CH_3Cl from salt marshes (site M2c) exceed those from freshwater wetlands⁸ by factors of about 750 and 500, respectively. Diurnal flux (ϕ) variations were modelled by the least-squares fitting of the daytime measurements to the function

$$\phi = a + b \left(1 + \cos \frac{t\pi}{8} \right)$$

where t is the difference in hours from the mean of the times of the midday temperature and sunlight maxima, and a and b are fitted constants; a represents the night-time baseline flux.

Averaged daily fluxes, determined by integrating the modelled fluxes over a 24-hour cycle at the three sites (M2a, M2b and M2c, respectively), are 5.0, 7.2 and 20 $\mu\text{mol m}^{-2} \text{d}^{-1}$ for CH_3Br and 75, 130 and 480 $\mu\text{mol m}^{-2} \text{d}^{-1}$ for CH_3Cl . During the course of the day, fluxes ranged from 0.5 to 2 times the daily average for CH_3Br and from 0.9 to 1.3 times the daily average for CH_3Cl . Spatial variability may be roughly assessed by comparing these averages from the upper-middle marsh. Assuming a lognormal distribution, the mean value of 9 $\mu\text{mol m}^{-2} \text{d}^{-1}$ for CH_3Br varies by a factor of 2.1, while the mean value of 170 $\mu\text{mol m}^{-2} \text{d}^{-1}$ for CH_3Cl varies by a factor of 2.6. Of the three upper-middle marsh sites sampled, the lowest flux emanated from the original site (M2a). Thus, the overall salt-marsh flux may be even higher than our seasonal sampling would suggest. Within a single vegetation zone, the spatial representation of fluxes can pose a larger uncertainty than the diurnal variability.

Certain reactions involving methyltransferase enzymes are known to produce CH_3Br and CH_3Cl through the *S*-adenosyl-L-methionine (SAM)-dependent methylation of bromide (Br^-) and chloride (Cl^-)^{9–11}; methyltransferase activity has also been identified in several salt-marsh plant species^{9,12}. Our observations of high

methyl halide emissions from salt marshes were not predicted, however, because the leaves of certain halophytic (salt-tolerant) plants were found to have relatively low methyltransferase activities compared to the leaves of Brassicaceae species¹². Possible explanations for the surprisingly large salt-marsh emissions involve the different species of halophytic plants under study, the influence of the whole plant on emissions, the interaction between the plant and associated microflora, and the high concentrations of readily available bromide and chloride ions.

The observed fluxes of CH_3Br and CH_3Cl are linearly correlated (Fig. 2), despite the major differences in salt-marsh zones, predominant vegetation, and month of the year. This correlation spans four orders of magnitude of fluxes, and points to a common mechanism of formation and release for these compounds. The average molar ratio of chloride ion to bromide ion in sea water is approximately 650, while the ratio of emissions of methyl chloride to methyl bromide from the salt marsh is 17 ± 14 . Therefore the mechanism appears to favour strongly the production of CH_3Br to that of CH_3Cl . This observation is generally consistent with the behaviour of the methyltransferases isolated from certain plants and wood-rot fungi^{9–11,13}, in which the enzymatic production of CH_3Br is kinetically favoured over that of CH_3Cl . The ratios of emissions of these compounds from phytoplankton¹⁴ and macroalgae¹⁵ incubated in sea water also show that the production of CH_3Br is favoured relative to CH_3Cl , when compared to the ratio of halide ion concentrations in sea water. A more quantitative comparison of the observed emission ratios from field studies to those predicted by enzyme kinetics is complicated by the possibility that chloride and bromide concentrations in seawater differ from the concentrations at the active site of the enzyme, due to selective uptake or exclusion of ions by the cell.

Global extrapolations indicate that salt marshes may contribute significantly towards balancing the known global budgets of CH_3Br and CH_3Cl . The global contribution of salt marshes to these source fluxes can be estimated if we make the following assumptions: that our seasonal flux-chamber sites are representative of their respective salt marshes; that the diurnal trend (adjusted for length of day) and spatial variability of fluxes observed in our upper-middle marsh study are representative of each salt-marsh vegetation zone and season of study; that our study sites are representative of salt marshes globally; and that the global salt marsh area is $0.38 \times 10^{12} \text{ m}^2$ (ref. 2). Accordingly, we estimate that salt marshes globally may be responsible for annual releases to the troposphere of 0.14 Gmol (14 Gg) of CH_3Br and 3.3 Gmol (170 Gg) of CH_3Cl (Gmol, 10^9 mol ; Gg, 10^9 g). The full range of estimates, based on the spatial variability of fluxes, is 0.08 – 0.31 Gmol (7 – 29 Gg) per year of CH_3Br and 1.3 – 8.6 Gmol (65 – 440 Gg) per year of CH_3Cl . Before this present work, the largest estimated natural terrestrial sources for these compounds were 7 Gg $\text{CH}_3\text{Br yr}^{-1}$ from *Brassica* plants¹⁶ and 160 Gg $\text{CH}_3\text{Cl yr}^{-1}$ from wood-rot fungi¹⁷. Consequently, salt marshes may constitute the largest natural terrestrial source of CH_3Br , and possibly of CH_3Cl , identified thus far.

Because our sampling is both temporally and spatially limited, global extrapolations are subject to large uncertainties. The estimate of salt-marsh area alone has an uncertainty of 50% or more². Furthermore, southern California salt marshes differ in structure and composition from other salt marshes. For example, salt marshes on the east coast of the United States are dominated by *Spartina alterniflora*², and at lower latitudes, salt marshes are replaced by mangrove forests which have greater productivity and higher-density above-ground biomass¹⁸. Fluxes are much higher during the growing season, which may be significant when interpreting the seasonal cycles of background atmospheric concentrations¹⁹. If high concentrations of halide ions are critical to the production of CH_3Br and CH_3Cl by plants, then halophytic plants of tropical mangrove and semi-arid environments may also release significant amounts of these compounds.

Methods

Sampling

The aluminium flux chamber consists of a collar, which is seated in the soil first, and a lid, which is placed on top to initiate the measurement period. Three whole air samples are drawn from the flux chamber into previously evacuated 6-litre fused-silica-lined stainless steel canisters at 15- or 20-minute intervals. Small electric fans circulate the chamber air before the last two samplings, and a chamber vent tube is opened during sampling to allow for pressure equilibration. Ambient temperature, chamber temperature, and atmospheric pressure are monitored, and ambient air samples are taken at the beginning and end of each flux chamber experiment.

Gas measurements

CH₃Br and CH₃Cl were measured by capillary column gas chromatography with cryogenic preconcentration on a Porapak Q capillary microtrap, separation on a PoraPLOT Q column, and oxygen-doped electron capture detection^{5,20}. Calibration curves are constructed with a suite of synthetic primary standards prepared in-house⁵ which span the range of the concentrations. CH₄ was measured by packed-column gas chromatography, with separation on a silica gel precolumn and a molecular sieve 5-Å column, and flame ionization detection²¹. CO₂ was measured by packed-column gas chromatography, with separation on a Porapak Q column, nickel catalyst reduction, and flame ionization detection. Flask sub-samples are measured twice on each of these instruments, with relative analytical precisions typically ranging from 0.2 to 2%.

Received 8 September; accepted 5 November 1999.

- Kurylo, M. J. & Rodriguez, J. M. in *Scientific Assessment of Ozone Depletion: 1998* (eds Albritton, D. L., Watson, R. T. & Aucamp, P. J.) (Report no. 44, World Meteorological Organization, Geneva, 1999).
- Woodwell, G. M., Rich, P. H. & Hall, C. A. S. in *Carbon and the Biosphere* (eds Woodwell, G. M. & Pecan, E. V.) 221–240 (Brookhaven Symposia in Biology, 24, US Atomic Energy Commission, Washington DC, 1973).
- Zedler, J. B. *The Ecology of Southern California Coastal Salt Marshes: A Community Profile* (FWS/OBS-81/54, US Fish & Wildlife Service, Biological Services Program, Washington DC, 1982).
- Livingston, G. P. & Hutchinson, G. L. in *Biogenic Trace Gases: Measuring Emissions from Soil and Water* (eds Matson, P. A. & Harriss, R. C.) 14–51 (Blackwell Science, Oxford, 1995).
- Miller, B. R. *Abundances and Trends of Atmospheric Chlorodifluoromethane and Bromomethane*. Thesis, Univ. California, San Diego (1998).
- Delaune, R. D., Smith, C. J. & Patrick, W. H. Methane release from gulf coast wetlands. *Tellus B* 35, 8–15 (1983).
- Bartlett, K. B. Methane flux from coastal salt marshes. *J. Geophys. Res.* 90, 5710–5720 (1985).
- Varner, R. K., Crill, P. M. & Talbot, R. W. Wetlands: a potentially significant source of atmospheric methyl bromide and methyl chloride. *Geophys. Res. Lett.* 26, 2433–2436 (1999).
- Wuostmaa, A. M. & Hager, L. P. Methyl chloride transferase: a carbocation route for biosynthesis of halomethanobolites. *Science* 249, 160–162 (1990).
- Attieh, J. M., Hanson, A. D. & Saini, H. S. Purification and characterization of a novel methyl-transferase responsible for biosynthesis of halomethanes and methanethiol in *Brassica oleracea*. *J. Biol. Chem.* 270, 9250–9257 (1995).
- Saxena, D., Aouad, S., Attieh, J. & Saini, H. Biochemical characterization of chloromethane emission from the wood-rotting fungus *Phellinus pomaceus*. *Appl. Environ. Microbiol.* 64, 2831–2835 (1998).
- Saini, H. S., Attieh, J. M. & Hanson, A. D. Biosynthesis of halomethanes and methanethiol by higher plants via a novel methyltransferase reaction. *Plant Cell Environ.* 18, 1027–1033 (1995).
- Harper, D. Halomethane from halide ion—a highly efficient fungal conversion of environmental significance. *Nature* 315, 55–57 (1985).
- Scarratt, M. G. & Moore, R. M. Production of methyl chloride and methyl bromide in laboratory cultures of marine phytoplankton. *Mar. Chem.* 54, 263–272 (1996).
- Manley, S. L. & Dastoor, M. N. Methyl iodide (CH₃I) production by kelp and associated microbes. *Mar. Biol.* 98, 477–482 (1988).
- Gan, J., Yates, S. R., Ohr, H. D. & Sims, J. J. Production of methyl bromide by terrestrial higher plants. *Geophys. Res. Lett.* 25, 3595–3598 (1998).
- Watling, R. & Harper, D. Chloromethane production by wood-rotting fungi and an estimate of the global flux to the atmosphere. *Mycol. Res.* 102, 769–787 (1998).
- Alongi, D. M. *Coastal Ecosystem Processes* (eds Kennish, M. J. & Lutz, P. L.) (CRC, Boca Raton, 1998).
- Wingenter, O. W., Wang, C. J.-L., Blake, D. R. & Rowland, F. S. Seasonal variation of tropospheric methyl bromide concentrations: Constraints on anthropogenic input. *Geophys. Res. Lett.* 25, 2797–2800 (1998).
- Miller, B. R., Huang, J., Weiss, R. F., Prinn, R. G. & Fraser, P. J. Atmospheric trend and lifetime of chlorodifluoromethane (HCFC-22) and the global tropospheric OH concentration. *J. Geophys. Res.* 103, 13237–13248 (1998).
- Steele, L. P. *et al.* The global distribution of methane in the troposphere. *J. Atmos. Chem.* 5, 125–171 (1987).

Acknowledgements

We thank J. Fessenden, M. Vollmer, G. Mortyn, D. Matross and T. Tanhua for field assistance; P. Salameh and S. Walker for programming support; D. Harper for comments; and the University of California Natural Reserve System (NRS) and the Joint Powers Authority of the San Dieguito River Park for fieldwork permits. This work was supported by the NRS, the Methyl Bromide Global Coalition, NASA's Upper Atmosphere Research Program, and the NSF Graduate Research Fellowship Program.

Correspondence and requests for materials should be addressed to R.C.R. (e-mail: rob@gaslab.ucsd.edu).

A strong source of methyl chloride to the atmosphere from tropical coastal land

Y. Yokouchi*, Y. Nojiri*, L. A. Barrie†, D. Toom-Saunty†, T. Machida*, Y. Inuzuka*, H. Akimoto‡, H.-J. Li*, Y. Fujinuma* & S. Aoki§

* National Institute for Environmental Studies, 16-2, Onogawa, Tsukuba, Ibaraki 305-0053, Japan

† Atmospheric Environment Service, 4095, Dufferin St., Toronto, Ontario M3H 5T4, Canada

‡ Research Center for Advanced Science and Technology, University of Tokyo, 4-6-1, Komaba, Meguro-ku, Tokyo 153-8904, Japan

§ Center for Atmospheric and Oceanic Studies, Tohoku University, Aoba, Aramaki, Aoba-ku, Sendai 980-0845, Japan

Methyl chloride (CH₃Cl), the most abundant halocarbon in the atmosphere, has received much attention as a natural source of chlorine atoms in the stratosphere^{1,2}. The annual global flux of CH₃Cl has been estimated to be around 3.5 Tg on the grounds that this must balance the loss through reaction with OH radicals (which gives a lifetime for atmospheric CH₃Cl of 1.5 yr)^{3–5}. The most likely main source of methyl chloride has been thought to be oceanic emission^{2,6–8}, with biomass burning the second largest source⁹. But recent seawater measurements¹⁰ indicate that oceanic fluxes cannot account for more than 12% of the estimated global flux of CH₃Cl, raising the question of where the remainder comes from. Here we report evidence of significant CH₃Cl emission from warm coastal land, particularly from tropical islands. This conclusion is based on a global monitoring study and spot measurements, which show enhancement of atmospheric CH₃Cl in the tropics, a close correlation between CH₃Cl concentrations and those of biogenic compounds emitted by terrestrial plants, and OH-linked seasonality of CH₃Cl concentrations in middle and high latitudes. A strong, equatorially located source of this nature would explain why the distribution of CH₃Cl is uniform between the Northern and Southern hemispheres, despite their differences in ocean and land area.

The seasonality of sources and sinks of atmospheric CH₃Cl at different latitudes was studied by periodical monitoring at Hateruma Island, Japan, a small island (12.5 km²) situated 250 km east of Taiwan in the subtropics (latitude 24.1°N, longitude 123.8°E); at Alert, Canada, in the Arctic (latitude 82.5°N, longitude 62.3°W); and over the northwest Pacific Ocean. Over the northwest Pacific Ocean, 7 samples were collected between 42°N and 54°N with an interval of 2° on board the cargo ship *Skaugran*, which sails regularly between Japan and Canada. Figure 1 shows the sampling locations as well as those for the other observations in this study. All samples were collected using evacuated stainless steel canisters, 6-litre fused-silica lined canisters (Silico-can, Restek), or 3-litre electro-chemical buffering canisters (ECB, Ultrafinish Technology), in both of which CH₃Cl was stable for more than six months (ref. 11). Samples were analysed using pre-concentration/capillary gas chromatography/mass spectrometry after being transported to the laboratory. Details of our analytical method have been published elsewhere^{11,12}. Data on CH₃Cl obtained from the three data sets for July 1996 to December 1998 are plotted in Fig. 2 to show seasonal variation; each year is represented by a different symbol (one data point from the 12 May 1998 sample from the northwest Pacific (761 p.p.t.) that was far beyond the average was omitted for the following discussion). The data sets from Alert and the northwest Pacific showed clear seasonal variation (Fig. 2a and b). Abundance and seasonal pattern, exhibiting a slight decrease of CH₃Cl in autumn, are very similar for the data sets from Alert and the northwest Pacific, and both data sets

can be fitted sinusoidally ($R = 0.78$ and 0.83 ; mean values, 500 and 527 p.p.t.v.; and peak to valley amplitudes of 96 and 95 p.p.t.v., for Alert and the northwest Pacific, respectively). Such seasonal change is consistent with the seasonally varying concentration of hydroxyl radicals that predominantly react with CH_3Cl , suggesting little seasonal effect from local sources of atmospheric CH_3Cl in those areas. No influence from strong phytoplankton blooming in early summer in the northwest Pacific was detected in the CH_3Cl variation there, which is consistent with the absence of correlation between the levels of CH_3Cl in the surface water and chlorophyll¹⁰. The average concentration over the northwest Pacific, which is almost the same as that reported for the Labrador Sea in similar latitudes¹⁰, could be taken as the representative mid-latitude concentration over the open ocean. The data set from subtropical Hateruma Island shows a different seasonal pattern in CH_3Cl (Fig. 2c). The CH_3Cl concentrations there were substantially higher than those at Alert in the Arctic and in the northwest Pacific especially in late spring and summer. The high concentration (mean, 600 p.p.t.v.), as well as higher variability of atmospheric CH_3Cl suggest a nearby large source of emission.

We observed an enhancement of CH_3Cl in air over subtropical islands at Okinawa Island during recent atmospheric investigations (2–21 August 1996 and 19 July–3 August 1997). Okinawa Island is mostly covered by sub-tropical forests, and sampling was done at Cape Hedo at the north end of the island. Part of the 1996 results has been published elsewhere¹²; concentrations as high as 1,500 p.p.t. of CH_3Cl were detected at Cape Hedo, Okinawa (latitude 26.9°N , longitude 128.3°E), during a calm night. High production of CH_3Cl in coastal waters surrounding the island was considered as a possible reason, since it was difficult to discriminate between coastal water sources and coastal land sources using only the CH_3Cl measurement at the site; also, marine sources were believed to be important at that time. However, it is now likely¹⁰ that CH_3Cl emission is not closely related to marine bioactivity, and significant emission of CH_3Cl cannot be expected even from coastal waters. Thus, a land source of CH_3Cl may be responsible for the increase. In the summer of 1997,

the diurnal variation of atmospheric organic compounds was observed intensively for five consecutive days (based on sampling every four hours in the period of 19–23 July 1997); CH_3Cl concentration was also found to increase on calm nights. Concurrent measurements of some other volatile compounds showed high correlation ($R = 0.92$) of CH_3Cl with α -pinene which is a very familiar biogenic compound emitted by many kinds of terrestrial plants^{13,14}. The observed time series of CH_3Cl concentration is shown in Fig. 3 as well as those of wind speed and wind direction measured at the same location. Significantly high levels of CH_3Cl were observed at 3:00 and 19:00 on 20 July when there was little wind ($0.5\text{--}1\text{ m s}^{-1}$) coming from the south (forested area), while no concentration higher than 600 p.p.t.v. was observed when the wind came from the east (the ocean). It is probable that elevated CH_3Cl concentrations observed at this island site were related to emission from the forested land marked by plant-derived α -pinene. Spot measurements were also made at Jakarta (latitude 6.2°S , longitude 106.8°E) and in a forest of Bandung (latitude 6.9°S , longitude 107.6°E) on tropical Java, Indonesia, during September 1998; these showed extremely high concentrations of CH_3Cl ranging from 1,000 to 2,100 p.p.t.v.

Regarding terrestrial sources of CH_3Cl , Harper¹⁵ demonstrated the biosynthesis and subsequent emission of CH_3Cl by wood-rot fungi, and it has been shown^{16,17} that leaf-disks of many terrestrial

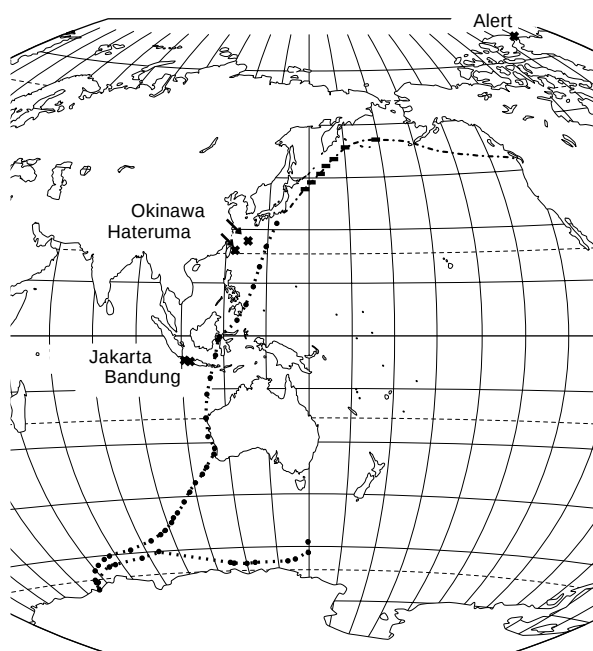


Figure 1 Sampling locations and cruise track of the 39th Japanese Antarctic Research Expedition. Shown are sampling points over the northwest Pacific (rectangles) and during the Shirase cruise (circles). The locations of named cities are indicated with crosses.

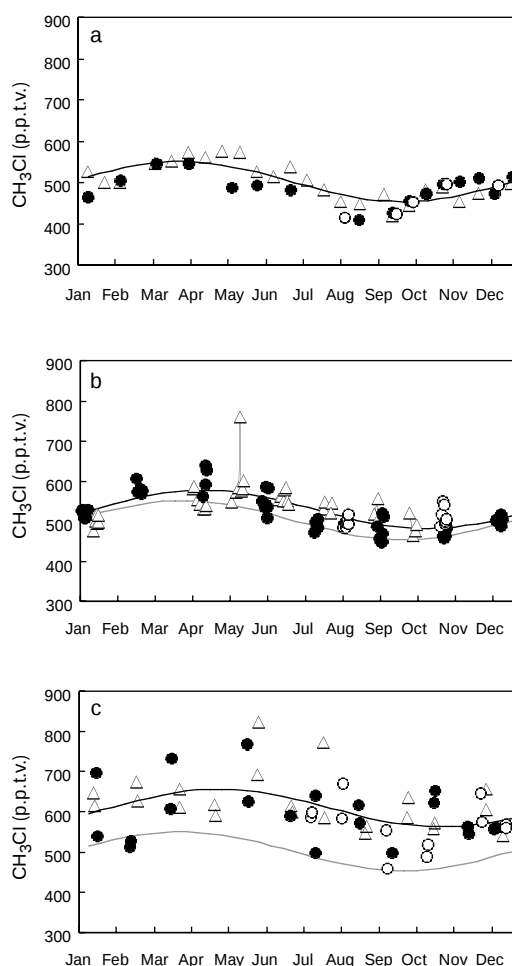


Figure 2 Seasonal change of atmospheric methyl chloride. Sampling locations are: **a**, Alert; **b**, northwest Pacific Ocean where 7 samples were collected between 42°N and 54°N ; **c**, Hateruma Island. Open circles, 1996 data; filled circles, 1997 data; open triangles, 1998 data. Thick line, sinusoidal regression for each dataset; thin line (**b**, **c**), sinusoidal regression for Alert data for comparison. p.p.t.v., parts per trillion by volume.

plants were able to synthesize methyl halides when placed in contact with halides in solution. If these processes were mediated through halide volatilization and elimination from plant tissues, it is probable that CH_3Cl emission would be more significant from coastal than inland plants. In coastal areas, where there is higher salinity (high content of Cl^-) in plants and soils, substantial quantities of CH_3Cl could be produced by biosynthesis. Such terrestrial emission is considered to be limited to coastal regions, as the CH_3Cl data from the northwest Pacific (in the same latitudinal region as North America and Eurasia) showed low values in this study. High temperatures in the tropics would enhance the biosynthetic production of CH_3Cl . High emission from tropical coastal lands is not only consistent with all the present observations, but can also account for previous discrepancies associated with an oceanic source only. These discrepancies are a uniform distribution between the Northern and Southern hemispheres despite the differing areas of ocean in the two hemispheres (intensive sources around the Equator would result in a uniform distribution between them); and elevation in the concentration of boundary layer CH_3Cl over the tropics observed by some researchers^{6,18} but not by others^{3,10} (the difference might depend on variations in the influence of air from tropical islands or coasts on these marine boundary layer samples).

The latitudinal distribution of CH_3Cl over the western Pacific–eastern Indian Ocean–Southern Ocean (31° N to 69° S (Antarctica)) was obtained during cruises of the 39th Japanese Antarctic Research

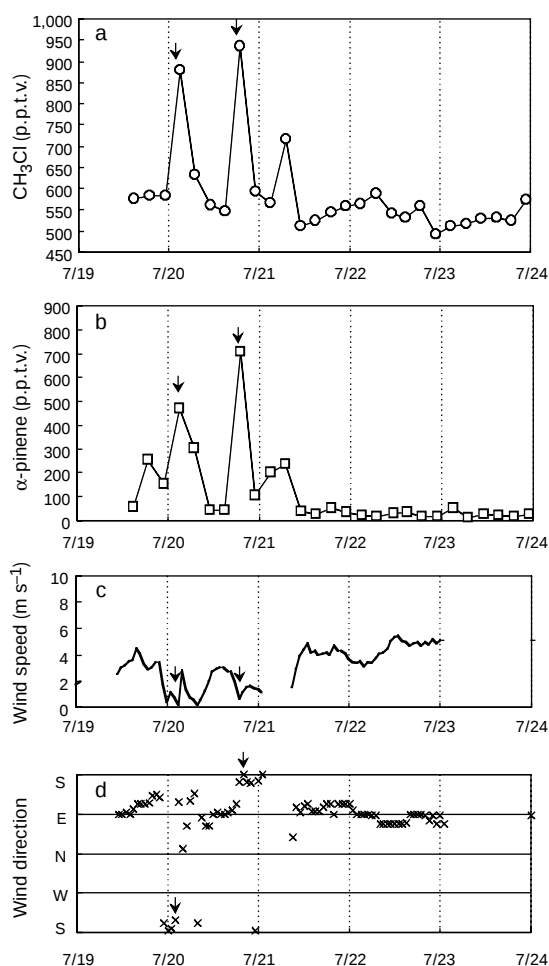


Figure 3 Measurements made at Cape Hedo of Okinawa Island. **a**, Variation of atmospheric CH_3Cl ; **b**, variation of atmospheric α -pinene; **c**, wind speed; **d**, wind direction. Dates given are 19–24 July 1997.

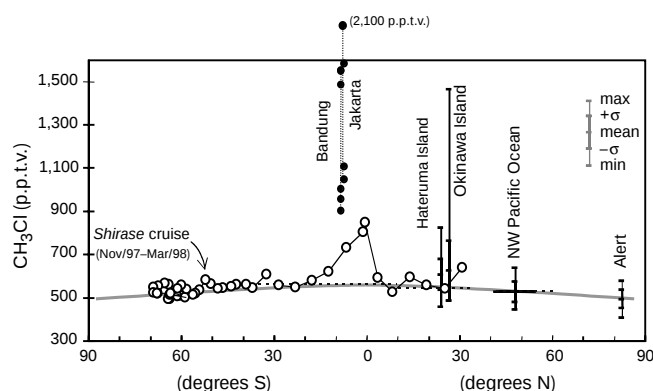


Figure 4 Latitudinal distribution of methyl chloride. Data were obtained from the RV *Shirase* cruise (open circles), monitoring at Alert, the northwest Pacific Ocean and Hateruma Island, summertime measurements at Okinawa Island, and spot measurements at Jakarta and Bandung. The broken line shows previously reported atmospheric CH_3Cl values for the northwest Atlantic and Pacific Ocean¹⁰. The grey line represents the most probable background distribution of CH_3Cl .

Expedition (RS *Shirase*; Leg 1, 14–27 November 1997; Leg 2, 3–16 December 1997; Leg 3, 15 February–21 March 1998) (Fig. 1). The *Shirase* data showed significant increases, up to 850 p.p.t., near the Equator. This equatorial increase would be partly due to the influence of high emission from the nearby tropical islands as shown above. However, there was a large-scale forest fire in Kalimantan and surrounding islands in 1997, and a considerable part of the CH_3Cl enhancement might be attributed to its influence. Apart from that enhancement, CH_3Cl levels from the *Shirase* cruises were constant, showing small decreases towards higher latitudes (1.2 p.p.t.v. decrease per degree from 23° S to 69° S). The baseline distribution of CH_3Cl was drawn (Fig. 4) using *Shirase* data (except those of the equatorial enhancement), mean values of the northwest Pacific and Alert data, and the lowest ends of subtropical data (which are expected to reflect oceanic conditions without influence from islands). The resulting distribution is very similar to the low- and mid-latitude oceanic data of Moore *et al.*¹⁰. Such a global distribution of CH_3Cl with slightly higher values near the Equator (~570 p.p.t.) than in the high Arctic and Antarctic (~500 p.p.t.) in Fig. 4, supports the concept of large CH_3Cl sources in the tropics. The latitudinal pattern with high CH_3Cl in the tropics is also similar to that obtained by Khalil and Rasmussen⁵, although their absolute concentration is higher than ours by 12%. They also observed higher concentrations of CH_3Cl at inland locations, suggesting continental sources; however no enhancement over the western North America continent was observed in a previous study⁶.

CH_3Cl has long been believed to be mainly emitted from the ocean, with some contribution from biomass burning. However, some studies^{5,10} have shown that the ocean cannot be a significant source and the present study suggests the greater importance of tropical islands as a source. For quantitative estimates of CH_3Cl emission from land sources, more measurements in coastal areas and over continents are required, as well as a modelling study on transport.

1. Fabian, P. in *The Handbook of Environmental Chemistry* (ed. Hutzinger, O.) Vol. 4/Part A, 23–51 (Springer, Berlin, 1986).
2. Graedel, T. E. & Keene, W. C. The tropospheric budget of reactive chlorine. *Glob. Biogeochem. Cycles* **9**, 47–77 (1995).
3. Koppmann, R. *et al.* Distribution of methylchloride, dichloromethane, trichloroethene and tetrachloroethene over the North and South Atlantic. *J. Geophys. Res.* **98**(D11), 20517–20526 (1993).
4. Logan, J. A. *et al.* Tropospheric chemistry: A global perspective. *J. Geophys. Res.* **86**, 7210–7254 (1981).
5. Khalil, M. A. K. & Rasmussen, R. A. Atmospheric methyl chloride. *Atmos. Environ.* **33**, 1305–1321 (1999).

6. Rasmussen, R. A. *et al.* Concentration distribution of methyl chloride in the atmosphere. *J. Geophys. Res.* **85**, 7350–7356 (1980).
7. Singh, H. B., Salas, L. J. & Stiles, R. E. Methyl halides in and over the eastern Pacific (40°N–32°S). *J. Geophys. Res.* **88**, 3684–3690 (1983).
8. Ciccone, R. J. Halogens in the atmosphere. *Rev. Geophys. Space Phys.* **19**, 123–139 (1981).
9. Blake, N. J. *et al.* Biomass burning emissions and vertical distribution of atmospheric methyl halides and other reduced carbon gases in the south Atlantic region. *J. Geophys. Res.* **101** (D19), 24151–24164 (1996).
10. Moore, R. M., Groszko, W. & Niven, S. J. Ocean-atmosphere exchange of methyl chloride: results from NW Atlantic and Pacific Ocean studies. *J. Geophys. Res.* **101**, 28529–28539 (1996).
11. Yokouchi, Y. *et al.* Isoprene in the marine boundary layer (Southeast Asian Sea, Eastern Indian Ocean, Southern Ocean): comparison with DMS and bromoform. *J. Geophys. Res.* **101** (D7), 8067–8076 (1999).
12. Li, H.-J., Yokouchi, Y. & Akimoto, H. Measurements of methyl halides in the marine atmosphere. *Atmos. Environ.* **33**, 1881–1887 (1999).
13. Zimmerman, P. R. Testing of hydrocarbon emissions from vegetation, leaf litter and aquatic surfaces and development of a methodology for compiling biogenic emission inventories. (Report EPA-450/4-79-004, US Environmental Protection Agency, Research Triangle Park, North Carolina, 1979).
14. Yokouchi, Y. *et al.* Determination of monoterpene hydrocarbons in the atmosphere. *J. Chromatogr.* **209**, 293–298 (1981).
15. Harper, D. B. Halomethane from halide ion—a highly efficient fungal conversion of environmental significance. *Nature* **315**, 55–57 (1985).
16. Attieh, J. M., Hanson, A. D. & Saini, H. S. Purification and characterization of a novel methyl-transferase responsible for biosynthesis of halomethanes and methanethiol in *Rassia oleracea*. *J. Biol. Chem.* **270**, 9250–9257 (1995).
17. Saini, H. S., Attieh, J. M. & Hanson, D. Biosynthesis of halomethanes and methanethiol by higher plants via a novel methyl transferase reaction. *Plant Cell Environ.* **18**, 1027–1033 (1995).
18. Atlas, E. *et al.* Alkyl nitrates, nonmethane hydrocarbons, and halocarbon gases over the equatorial Pacific Ocean during saga 3. *J. Geophys. Res.* **98**, 16933–16947 (1993).

Acknowledgements

We thank staff at the Atmospheric Environment Service station for assistance in sample collection at Alert. We are grateful to the National Institute of Polar Research, Japan for the opportunity to participate in the 39th Japanese Antarctic Research Expedition. We also thank the staff of the Global Environmental Forum Foundation for their assistance with sample collection at Hateruma and with the *Skaugran* cruise. In addition, we also thank the staff of the Environmental Management Center (Indonesia) for their assistance in the field study at Jakarta and Bandung. This work was supported in part by the Science and Technology Agency of Japan, by the Global Environment Fund (Environmental Agency of Japan), and by the Atmospheric Environment Service of Canada.

Correspondence and requests for materials should be addressed to Y.Y. (e-mail: yokouchi@nies.go.jp).

Halocarbons produced by natural oxidation processes during degradation of organic matter

F. Keppler, R. Eiden, V. Niedan, J. Pracht & H. F. Schöler

Institute of Environmental Geochemistry, Heidelberg University, Im Neuenheimer Feld 236, D-69120 Heidelberg, Germany

Volatile halogenated organic compounds (VHOC) play an important role in atmospheric chemical processes—contributing, for example, to stratospheric ozone depletion^{1–4}. For anthropogenic VHOC whose sources are well known⁵, the global atmospheric input can be estimated from industrial production data. Halogenated compounds of natural origin can also contribute significantly to the levels of VHOC in the atmosphere⁶. The oceans have been implicated as one of the main natural sources^{7–10}, where organisms such as macroalgae and microalgae can release large quantities of VHOC to the atmosphere^{11,12}. Some terrestrial sources have also been identified, such as wood-rotting fungi¹³, biomass burning¹⁴ and volcanic emissions¹⁵. Here we report the identification of a different terrestrial source of naturally occurring VHOC. We find that, in soils and sediments, halide ions can be alkylated

during the oxidation of organic matter by an electron acceptor such as Fe(III): sunlight or microbial mediation are not required for these reactions. When the available halide ion is chloride, the reaction products are CH₃Cl, C₂H₅Cl, C₃H₇Cl and C₄H₉Cl. (The corresponding alkyl bromides or alkyl iodides are produced when bromide or iodide are present.) Such abiotic processes could make a significant contribution to the budget of the important atmospheric compounds CH₃Cl, CH₃Br and CH₃I.

In a study of VHOC in surface peat-bog waters containing suspended organic matter, we found that besides the well-known trichloromethane which is attributed to an enzymatic halogenation of organic compounds¹⁸, significant amounts of other halocarbons such as iodomethane, iodoethane, iodopropane and iodobutane were released from organic-rich waters. The water samples displayed conspicuous concentrations of Fe(III) (up to 0.1 mmol l⁻¹), a high iodide content and a pH value of 4.2. Investigations of three different soil samples from anthropogenic unpolluted areas in Western Patagonia/Chile and Hawaii (Table 1) showed partially higher concentrations of CH₃Cl, C₂H₅Cl and CH₃Br in addition to the alkyl iodides. Isolation and identification of VHOC were performed by a headspace gas chromatography technique with electron capture and mass spectrometry detection.

To differentiate between biological or geochemical processes, a part of each sample was dried at 105 °C in order to destroy biological active material, such as microorganisms or enzymes, while the other part was freeze-dried. Both samples, when resuspended in distilled water, showed a similar methyl halide distribution to those we found in the natural untreated samples.

We considered that in soils there should exist an abiotic reaction mechanism that forms alkyl halides. The thermodynamically labile organic matter is oxidized and the redox partner (for example, Fe(III)) is reduced (to Fe(II)). Phenolic moieties of the natural organic matter containing methoxy groups might be oxidized while Fe(III) is reduced. During this process halides are methylated, and the methyl halides formed represent degradation products of oxidized organic matter (Fig. 1). Ethoxylated and propoxylated phenolic structures could be responsible for the occurrence of ethyl and propyl halides.

To verify this hypothesis the following experiments were initiated: (1) three different organic-rich soils (Table 1) from the natural reserve Rotwasser (Rotw.) in the Hessian Odenwald (Germany) were collected and tested for VHOC emission in relation to oxidant concentration and halide ion content; (2) for comparison, the same parameters were measured using 2-methoxyphenol (guaiacol), which represents a natural monomeric constituent of humus.

We suspended dried soil samples in sterile distilled water containing Fe(III) and halide ions. Fe(III) was readily reduced to Fe(II) depending on the organic carbon content of the soil samples (Fig. 2a). The higher the organic carbon content of the soil sample the more Fe(III) was reduced during the same period.

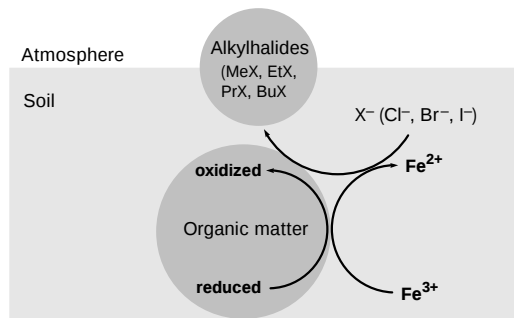


Figure 1 Model for alkyl halide formation by the reaction of Fe(III) and organic matter in the presence of halide ions.

Table 1 Organic carbon, pH, Fe, halogens and halogen ratio in different soil samples

Soil	C _{org}	pH	Fe	Cl	Br	I	Cl:Br:I (mol)
	(% dry wt)		(% dry wt)				
Grassland soil A _h (Rotw.)	6.6	4.96	0.70	187	<5.7	<3.7	
Coniferous forest soil A _h (Rotw.)	14.8	3.01	0.39	233	15.8	6.6	123:2.3:1
Suspended matter from peat water (Rotw.)	28.6	4.18	2.95	676	45.9	23.3	100:3:1
Forest soil O _h /A _h (Hawaii I)	26.3	5.6	1.7	195	33.3	37.3	15:1.4:1
Forest soil A _h (Hawaii II)	13.5	4.2	15.4	236	101	236	2.8:0.6:1
Grassland soil O _h /A _h (Patagonia)	46.9	4.10	0.26	2,900	260	32.9	320:12.5:1

Subscript h indicates soil horizon; C_{org} is organic carbon content; Rotw. indicates Rotwasser, Germany (see text); wt, weight.

VHOC release rates were measured during the reduction of Fe(III) in the presence of halide ions. We identified four different alkyl halides with decreasing release rates in the following order: halomethane > haloethane > 1-halopropane ≥ 1-halobutane. To quantify the alkyl halides we used iodide as the halide source. As a result, iodomethane, iodoethane, 1-iodopropane and 1-iodobutane

were formed in the proportion 19:5:1.5:1 (Fig. 2b). An almost similar proportion (17:4.5:0.5:1) was obtained from untreated peat-bog water.

The influence of Fe(III) concentration on the alkyl halide formation can be inferred from Fig. 3a. When higher amounts of Fe(III) were added to the soil, the alkyl halide emission increased. There was a little production of alkyl halides and Fe(II) even if no Fe(III) was added to the soil. This is probably caused by the presence of Fe(III) as a mineral phase in the form of iron oxides and oxyhydroxides in the soil samples (0.7–3% of dry weight) which could be reductively dissolved by natural organic compounds.

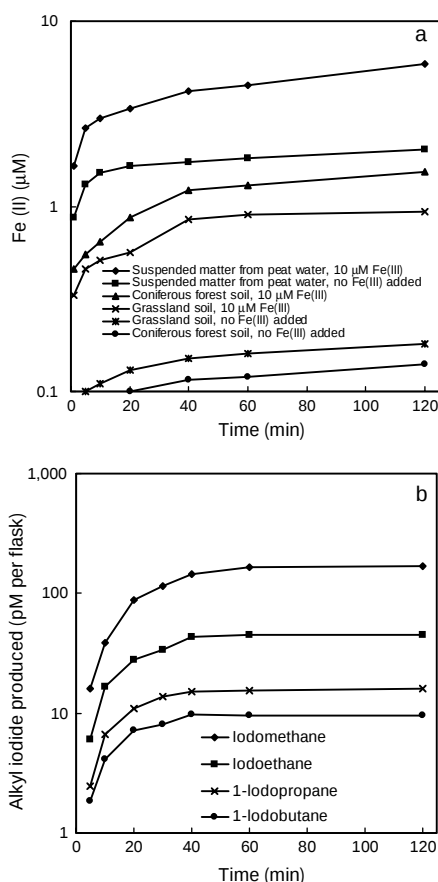


Figure 2 Abiotic reduction of Fe(III) in soils and the production of alkyl iodides in the presence of iodide. **a**, Abiotic reduction of Fe(III) to Fe(II) by different organic-rich soils. 100 mg soil sample (dried at 105 °C) was suspended in 10 ml bidistilled water containing 10 μM dissolved Fe(III) as electron acceptor. The pH in the medium was 3.0 and temperature was 30 °C. Subsamples were filtered (0.45 μm pore filter) and Fe(II) was photometrically analysed as ferrous iron-phenanthroline complex. Three to six replicates were measured for each sample ($n = 3-6$). The standard deviation (s.d.) was in the range 7–31% error (σ). **b**, Alkyl iodide formation of a grassland soil (Rotw.) in the presence of Fe(III) and iodide. 100 mg grassland soil (dried at 105 °C) was added to 10 ml bidistilled water in a 20-ml headspace flask containing 10 μM dissolved Fe(III). Before the flask was sealed the medium (pH 3.0) was supplemented with 100 μM KI. After shaking at a temperature of 30 °C the volatile halogenated organic compounds (VHOC) in the headspace were determined by gas chromatography (GC) with an electron capture detector (ECD) ($n = 4-6$; s.d. 4–25% error).

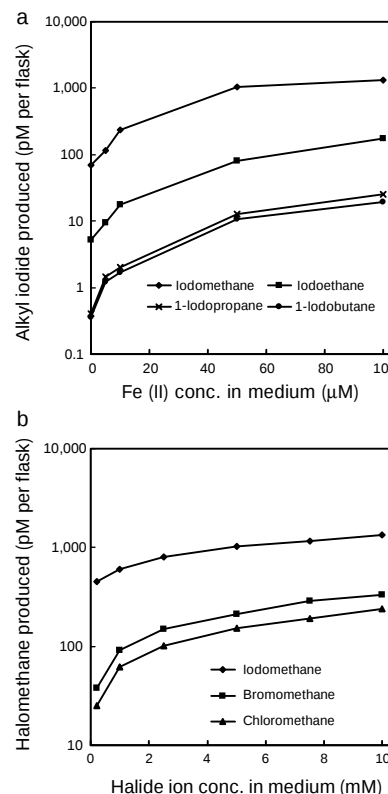
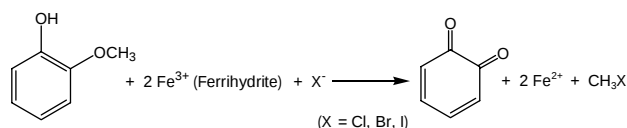


Figure 3 Influence of Fe(III) and halide ions on the formation of alkyl halides. **a**, Alkyl iodide production from a coniferous forest soil (Rotw.) by using Fe(III) as electron acceptor. Coniferous forest soil (500 mg, dried at 105 °C) was suspended in 10 ml bidistilled water. In 20-ml headspace flasks. The medium (pH 3.0) was supplemented with KI (100 μM) and 5–100 μM Fe(III). The flasks were sealed and shaken for 1 h at a temperature of 30 °C. VHOC concentrations in the headspace were measured by GC–ECD ($n = 4-5$; s.d. 5–44% error). **b**, Effect of halide ion concentration on yield of corresponding halomethane production with a grassland soil (Rotw., from Rotwasser, Germany, see text) as organic matter source. 1,000 mg grassland soil (freeze-dried) was added to 10 ml bidistilled water in a 20-ml headspace flask. The medium (pH 5.0) was supplemented with 1–10 mM KCl, KBr or KI before the flask was sealed. After shaking for 60 min at a temperature of 30 °C, halomethane concentration in the headspace was monitored by GC–ECD ($n = 4-6$; s.d. 7–25% error).

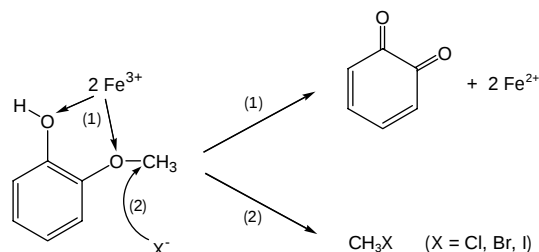
Figure 3b illustrates the effect of different halide concentrations on the yield of volatile halomethanes with a grassland soil sample (Rotw.) as the organic matter source. Halomethane production increased with increasing halide ion concentration. When equimolar concentrations (1 mM) of chloride, bromide and iodide were present, the preferred halide ion was iodide, the respective halomethanes being formed in the proportion of 1:1.5:10. We applied this experimentally determined proportion to the natural molar halogen ratios of the Patagonian soil (Cl:Br:I; 320:12.5:1), the Hawaiian soils I (15:1.4:1) and II (2.8:0.6:1), to obtain $\text{CH}_3\text{Cl}:\text{CH}_3\text{Br}:\text{CH}_3\text{I}$ ratios of 32:1.8:1, 7:1:5 and 3:1:11, respectively. Very similar ratios of 48:1.9:1, 5.5:1:2.9 and 3.5:1:8.7 were measured from these natural untreated samples. Our results suggest that a fraction of the soil halogen content is methylated by natural oxidation processes and that this fraction increases in the sequence $\text{Cl} < \text{Br} < \text{I}$.

The formation of methyl halides by a model reaction of guaiacol with ferrihydrite ($5\text{Fe}_2\text{O}_3 \cdot 9\text{H}_2\text{O}$) and halide is presented in Fig. 4. These findings are in agreement with the data we presented from natural soils. Ferrihydrite, the most common initial precipitate of

iron oxyhydroxides in nature, was used to oxidize guaiacol. Methyl halide, Fe(II) and *o*-quinone have been identified as reaction products.



There was no production of halomethanes if ferrihydrite or halide was absent. We assume that methyl halides are produced in an almost synchronous reaction scheme: (1) the oxidation of guaiacol by ferrihydrite and (2) nucleophilic substitution of the methyl group by halide.



The low molar ratio of halomethane to Fe(II) (Fig. 4a) indicates that further oxidation products of guaiacol were produced. The formation of a methoxyphenol dimer as a result of oxidative oligomerization of clays has been reported¹⁷. However, our reaction model describes a chemical pathway leading from guaiacol to methyl halide. This reaction is specific only for the methyl halides but in additional experiments with ethoxyphenol and propoxyphenol the corresponding ethyl halides and propyl halides were found.

Soil is a highly dynamic and complex system where a multitude of biological, bio-mediated and abiotic processes interact. In this study we have shown that in soils and sediments there exists an abiotic process, hitherto unknown, forming alkyl halides, which depends on organic-matter content, halide-ion and oxidant concentrations. These findings may have important implications for the global production of VHOC. For example, it is known that: (1) the reduction of insoluble Fe(III) oxides and oxyhydroxides is one of the most significant geochemical processes that takes place in the sedimentary environment¹⁸; (2) many terrestrial soils exhibit high salinities (up to 1% of dry weight); (3) worldwide 1,500–2,200 Gt of organic carbon is stored as humus in the soil layers¹⁹, and alkoxy-phenolic structures, especially methoxylated phenolic structures, are monomeric constituents of humic substances. Nevertheless, the magnitude of VHOC production in soil is difficult to estimate. In our model experiments, a maximum of 1 of 30,000 carbon atoms was liberated from the soil as VHOC within one hour. We cannot use this relationship to estimate a global emission rate, but our results do show that the terrestrial ecosystem has an enormous potential to release CH_3Cl , CH_3Br and CH_3I into the atmosphere.

The described processes will depend on many factors, including the ambient environmental conditions (such as moisture, temperature and soil acidity) and the type of organic material within the soil. Additional studies and associated field measurements are required in order to calculate the release rates of VHOC from soils and sediments, and to distinguish further between biogenic and abiotic production. □

Received 2 September; accepted 14 December 1999.

- Crutzen, P. J. & Arnold, F. Nitric acid clouds formation in the cold Antarctic stratosphere: a major cause for the springtime 'ozone hole'. *Nature* **324**, 651–655 (1986).
- Solomon, S. Progress towards a quantitative understanding of Antarctic ozone depletion. *Nature* **347**, 347–354 (1990).

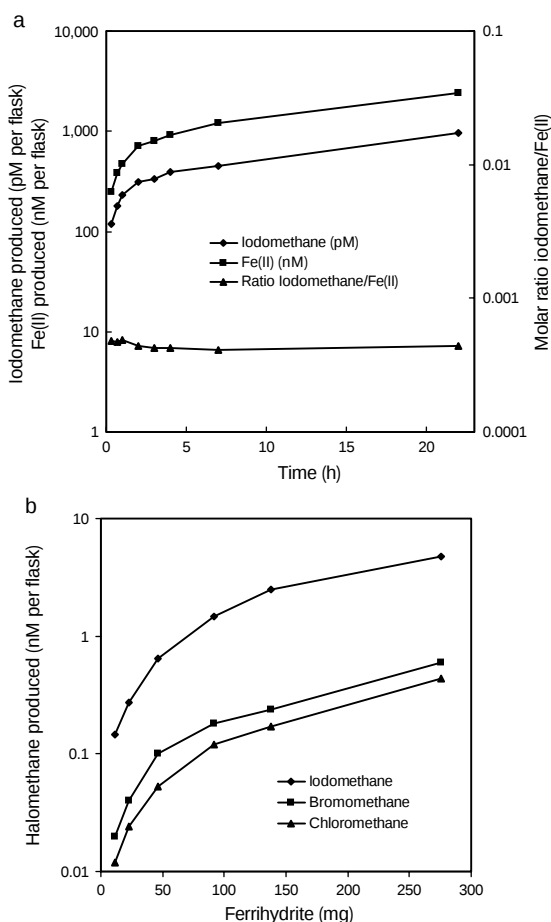


Figure 4 Data supporting the abiotic production of methyl halides by the interaction of guaiacol with ferrihydrite and halide. **a**, Reductive dissolution of ferrihydrite by guaiacol (2-methoxyphenol) and the related iodomethane formation. 20 mg ferrihydrite and 10 mM KI were suspended in 10 ml bidistilled water in a 20-ml headspace flask under argon atmosphere. The pH of the medium was adjusted with 1 M HNO_3 to a value of 4.4. Guaiacol (2 mM) was added and the flask was sealed. Subsamples were taken for Fe(II) and iodomethane detection ($n = 3-6$; s.d. 9–22% error). **b**, Influence of ferrihydrite on yield of halomethane production. 10 mM KCl, KBr or KI and 5–275 mg ferrihydrite were added to 10 ml bidistilled water in a 20-ml headspace flask. The pH of the medium was adjusted with 1 M HNO_3 to a value of 4.4. Before the flask was sealed the medium was supplemented with 2 mM guaiacol. After a reaction time of 60 min methyl halide was monitored ($n = 3-6$; s.d. 4–32% error).

3. Anderson, J. G., Toohey, D. W. & Brune, W. H. Free radicals within the Antarctic vortex: The role of CFCs in the Antarctic ozone loss. *Science* **251**, 39–46 (1991).
4. Wever, R. Ozone destruction by algae in the Arctic atmosphere. *Science* **355**, 501 (1988).
5. Butler, J. H. *et al.* A record of atmospheric halocarbons during the twentieth century from polar firm air. *Nature* **399**, 749–755 (1999).
6. Gribble, G. W. Naturally occurring organohalogen compounds—a survey. *J. Natural Prod.* **55**, 1353–1395 (1992).
7. Lovelock, J. E. Natural halocarbons in the air and in the sea. *Nature* **256**, 193–194 (1975).
8. Singh, H. B., Salas, L. J. & Stiles, R. E. Methylhalides in and over the Eastern Pacific (40° N–32° S). *J. Geophys. Res.* **88**, 3684–3690 (1983).
9. Class, T. & Ballschmiter, K. Sources and distribution of bromo- and bromochloromethanes in marine air and surface water of the Atlantic Ocean. *J. Atmos. Chem.* **6**, 35–46 (1988).
10. Moore, R. M., Groszko, W. & Niven, S. J. Ocean-atmosphere exchange of methyl chloride: Results from NW Atlantic and Pacific Ocean studies. *J. Geophys. Res.* **101**, 28529–28538 (1988).
11. Sturges, W. T., Sullivan, C. W., Schnell, R. C., Heidt, L. E. & Pollack, W. H. Bromoalkane production by Antarctic ice algae. *Tellus B* **45**, 120–126 (1993).
12. Laturnus, F. & Adams, F. C. Methylhalides from Antarctic macroalgae. *Geophys. Res. Lett.* **25**, 773–776 (1998).
13. Harper, D. B. Halomethane from halide ion—a highly efficient fungal conversion of environmental significance. *Nature* **315**, 55–57 (1985).
14. Andreae, M. O. *et al.* Methylhalide emissions from savanna fires in southern Africa. *J. Geophys. Res.* **101**, 23603–23613 (1996).
15. Rasmussen, R. A., Rasmussen, L. E., Khalil, M. A. K. & Dalluge, R. W. Concentration of methyl chloride in the atmosphere. *J. Geophys. Res.* **85**, 7350–7356 (1980).
16. Hoekstra, E. J., Lassen, P., van Leeuwen, J. G. E., De Leer, E. W. B. & Carlsen, L. in *Naturally Produced Organohalogen* (eds Grimvall, A. & De Leer, E. W. B.) 149–158 (Kluwer Academic, Dordrecht, 1995).
17. Desjardins, S., Landry, J. A. & Farant, J. P. Effects of water and pH on the oxidative oligomerization of chloro and methoxyphenol by a montmorillonite clay. *J. Soil Contamin.* **8**, 175–195 (1999).
18. Lovley, D. R. Dissimilatory Fe(III) and Mn(IV) reduction. *Microbiol. Rev.* **55**, 259–287 (1991).
19. Post, W. M., Emanuel, W. R., Zinke, P. J. & Stangenberger, A. L. Soil carbon pools and world life zones. *Nature* **298**, 156–159 (1982).

Acknowledgements

We thank D. Schlösser and K. Kratz for the instrumental neutron activation analysis (INAA) measurements and I. Fahimi and L. Warr for reviewing the manuscript.

Correspondence and requests for materials should be addressed to E.K. (e-mail: fkeppeler@ix.urz.uni-heidelberg.de).

Annual fluxes of carbon from deforestation and regrowth in the Brazilian Amazon

R. A. Houghton*, D. L. Skole†, Carlos A. Nobre‡, J. L. Hackler*, K. T. Lawrence* & W. H. Chomentowski†

* Woods Hole Research Center, PO Box 296, Woods Hole, Massachusetts 02543, USA

† Department of Geography, Michigan State University, East Lansing, Michigan 48824, USA

‡ Instituto Nacional de Pesquisas Espaciais, Caixa Postal 515, São José dos Campos, SP, CEP 12201-970, Brazil

The distribution of sources and sinks of carbon among the world's ecosystems is uncertain. Some analyses show northern mid-latitude lands to be a large sink, whereas the tropics are a net source¹; other analyses show the tropics to be nearly neutral, whereas northern mid-latitudes are a small sink^{2,3}. Here we show that the annual flux of carbon from deforestation and abandonment of agricultural lands in the Brazilian Amazon was a source of about 0.2 Pg C yr⁻¹ over the period 1989–1998 (1 Pg is 10¹⁵ g). This estimate is based on annual rates of deforestation and spatially detailed estimates of deforestation, regrowing forests and biomass. Logging may add another 5–10% to this estimate⁴, and fires may double the magnitude of the source in years following a drought⁴. The annual source of carbon from land-use change and fire approximately offsets the sink calculated for natural ecosystems in the region^{5,6}. Thus this large area of tropical forest is

nearly balanced with respect to carbon, but has an interannual variability of ± 0.2 PgC yr⁻¹.

We determined the annual flux of carbon with a 'bookkeeping' model^{7,8} that tracks the annual emission and uptake of carbon that follow the clearing of forest for agriculture and the regrowth of secondary forests on abandoned agricultural land. Changes in carbon include (1) the immediate loss of carbon to the atmosphere from plant material burned at the time of clearing, (2) the slower release of carbon from decay of dead plant material left on site (slash) and removed for wood products, and (3) the accumulation of carbon during forest growth. Changes in soil carbon were not included in this analysis, as they are small relative to the changes in biomass and are inconsistent in direction^{9–12}.

We used two estimates of deforestation, three estimates of biomass and two estimates of the rate of decay of organic matter to calculate a range of net carbon emissions attributable to land-use change. The first estimate of deforestation was obtained from the Brazilian Space Agency (INPE), where data from the Landsat satellite are delineated manually for each state to determine both annual rates of deforestation and cumulative areas deforested for each year between 1988 and 1998 (except 1993). The annual and cumulative data are not entirely consistent, and we used the cumulative areas deforested to calculate annual rates of change (Table 1). INPE also determined the area deforested in 1978; before 1960 rates of deforestation were negligible¹³.

The second estimate of deforestation was based on a map of land cover derived from classification of 1986 Landsat multi-spectral scanner data (Fig. 1). Areas classified as deforested in 1986 were consistently lower than INPE's 1988 estimate of deforested area. Because the dates were different, we interpolated a rate for 1988 based on maps of land cover derived from 1986 and 1992 Landsat data. The interpolated area deforested in 1988 was still about 25% lower than INPE's estimate, although the actual percentage varied among states (Fig. 2). We used this lower estimate for a second estimate of deforestation, varying it annually in proportion to the rates from INPE.

According to the Landsat-derived land-cover classification, about 30% of the deforested area was in secondary forest in 1986—presumably as a result of the abandonment of agricultural land^{14–17}. The percentage varied from 5% in Goiás to 65% in Maranhão. As we lacked data to suggest temporal trends in abandonment, we assumed that cleared lands were abandoned annually at the rate defined by the ratio of secondary forest to deforested area in 1986.

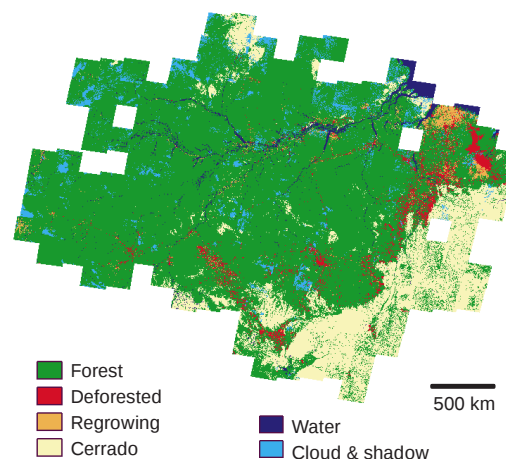


Figure 1 Land cover in Brazilian Amazonia as of 1986, based on a classification of Landsat MSS data. The classification identifies seven classes of land cover: forest, deforested land, regrowing forest, water, clouds, cloud shadow and *cerrado* (savanna). Here data for cloud and cloud shadow are grouped together.

Table 1 Rates of deforestation in the states of Brazilian Amazonia

Years	Rate of deforestation (10^6 ha yr^{-1})									
	78–88	88–89	89–90	90–91	91–92	92–94	94–95	95–96	96–97	97–98
Acre	0.064	0.090	0.050	0.040	0.040	0.048	0.124	0.044	0.036	0.057
Amapa	0.006	0.020	0.030	0.040	0.004	0.000	0.005	0.000	0.002	0.002
Amazonas	0.180	0.200	0.050	0.100	0.080	0.037	0.189	0.080	0.059	0.092
Maranhao	0.269	0.150	0.110	0.070	0.114	0.037	0.178	0.158	0.041	0.105
Mato Grosso	0.515	0.810	0.400	0.290	0.467	0.622	0.854	0.699	0.527	0.581
Para	0.751	0.780	0.490	0.380	0.379	0.428	0.865	0.713	0.414	0.556
Rondonia	0.258	0.180	0.170	0.110	0.226	0.260	0.410	0.250	0.199	0.239
Roraima	0.026	0.090	0.020	0.040	0.028	0.024	0.016	0.024	0.018	0.016
Tocantins	0.184	0.070	0.060	0.050	0.041	0.033	0.067	0.034	0.027	0.036
Brazilian Amazonia	2.253	2.390	1.380	1.120	1.379	1.490	2.708	2.001	1.323	1.684

Data provided by INPE.

Thus, for example, the rate of abandonment in Maranhao each year was equivalent to 65% of the rate of deforestation.

We constructed three estimates of forest biomass. The first and second were based on Projeto RADAMBRASIL (Brazilian forest inventories) of stem-wood volumes on thousands of plots between 1973 and 1983¹⁸. We converted these volumes to total biomass with equations from refs 19 and 20, including an additional 20% for below-ground biomass^{19,21}. Total biomass was assumed to be 50% carbon. The intersection of this map of total biomass with the 1986 land-cover map (Fig. 1) defined the number of deforested pixels in each biomass class. We defined five broad classes of biomass (<100, 100–130, 130–160, 160–190 and >190 Mg C ha^{-1}), and calculated the area-weighted mean biomass for each class in each state on the basis of the classes deforested as of 1986 (Fig. 3). For this estimate, the classes of biomass ranged from 66 to 277 Mg C ha^{-1} .

A second estimate of biomass was based on a conversion of RADAMBRASIL stem-wood volumes to biomass²⁴. This conversion yielded an Amazonia-wide estimate 60% higher than one based on equations from ref. 18. We assumed that the larger estimate for the entire region applied to each cell. Thus the spatial distribution of biomass was equivalent to the first estimate, but each cell was 60% higher (Fig. 3).

A third estimate of biomass was based on 56 sites where live above-ground biomass was determined either indirectly (from large-area surveys of stem-wood volumes) or directly (from harvest of trees on plots between 0.2 and 2 ha). For sites where dead above-ground biomass or below-ground biomass was not measured, we increased live above-ground biomass by 10% (average of 20 sites) and 24% (ref. 21), respectively, to obtain total biomass. Multiple

measurements at the same location were averaged so that each site was represented by a single value. These values were distributed over Amazonia with a simple two-dimensional interpolation (Fig. 3).

With deforestation, the fractions of initial forest biomass burned, left as slash, removed for products and converted to 'elemental carbon' through burning were 0.2, 0.7, 0.08 and 0.02 (refs 23, 24), respectively. We assumed rates of decay as follows: woody material removed from the site (wood products) decayed with a rate constant

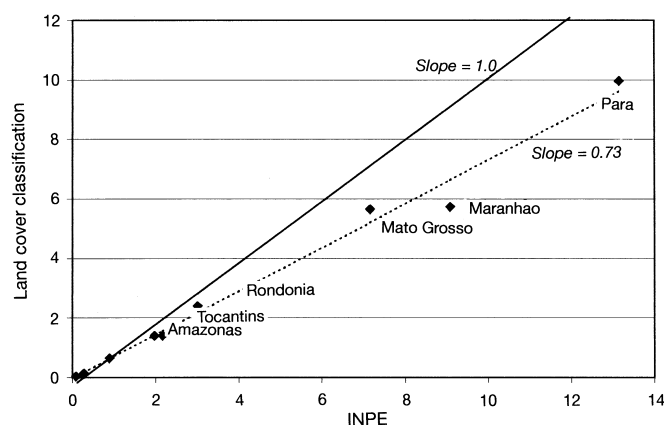


Figure 2 Comparison of two sets of estimates of the areas deforested by 1988 (10^6 ha) in each state of Brazilian Amazonia. 'INPE' indicates estimates from the Brazilian Space Agency, INPE. The rates obtained from a land-cover classification are based on an interpolation of 1986 and 1992 data.

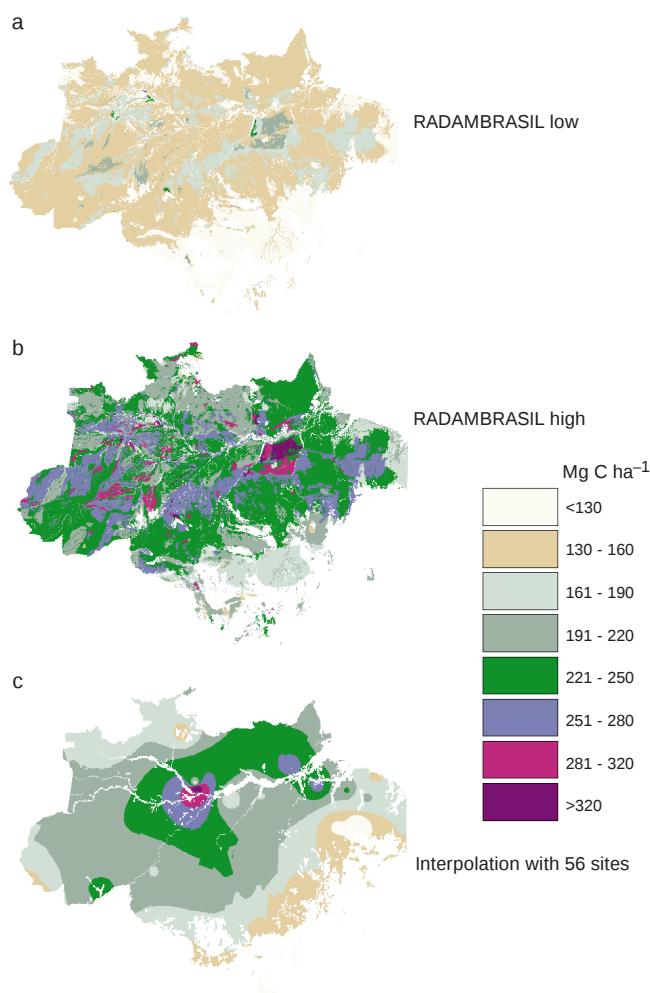


Figure 3 The spatial distribution of biomass in Brazilian Amazonia. **a**, RADAMBRASIL wood volumes converted to biomass with equations from refs 18, 19 (low estimate). **b**, RADAMBRASIL volumes converted to biomass (from ref. 21) (high estimate). **c**, Biomass interpolated from 56 sites (medium estimate).

Table 2 Eight combinations of data used to calculate a range of estimates of carbon flux

	Deforestation	Biomass	Decay rate	Carbon released 1989–1998	Carbon still to be released from slash products accumulated by 1998	Carbon to be accumulated in regrowth
1.	INPE	High	Low	2.64	3.03	3.12
2.	INPE	Medium	Low	2.14	2.46	1.32
3.	INPE	Medium	High	2.37	1.05	1.32
4.	INPE	Low	Low	1.63	1.87	1.00
5.	Land cover	High	Low	1.66	2.12	1.44
6.	Land cover	Medium	Low	1.34	1.72	1.17
7.	Land cover	Medium	High	1.59	0.77	1.17
8.	Land cover	Low	Low	1.02	1.31	0.88
			Mean	1.80	1.79	1.43

Shown are total amount of carbon (Pg C) released from deforestation and agricultural abandonment in Brazilian Amazonia over the period 1989–1998, carbon still to be released from slash and products remaining in 1998, and the total amount of carbon to be accumulated in forest growth if secondary forests remain undisturbed over the next 75 years.

of 0.1 yr^{-1} ; elemental carbon decayed at 0.001 yr^{-1} . Dead wood left on site decayed exponentially with one of two decay constants: 0.1 yr^{-1} for a 'best estimate', and 0.4 yr^{-1} to test the model's sensitivity to this parameter. Both values are within the range reported for 25 forests in Venezuela²⁵.

Abandoned agricultural lands reverted to forests that accumulated carbon at rates proportional to initial forest biomass. Rates ranged from about $1.5 \text{ Mg C ha}^{-1} \text{ yr}^{-1}$ in forests initially with biomass $< 100 \text{ Mg C ha}^{-1}$, to about $5.5 \text{ Mg C ha}^{-1} \text{ yr}^{-1}$ for forests initially $> 190 \text{ Mg C ha}^{-1}$. Growth of biomass is rapid in young forests, and declines with age²⁶. In this analysis, forests recovered 70% of their original biomass in 25 years and the remaining 30% over the next 50 years. The assumption that forests are fully regrown after as little as 75 years is probably not valid, but this analysis was largely concerned with human-induced changes over the past 10–40 years.

The two rates of deforestation described above, when used with the three estimates of biomass and two rates of decay, yielded eight estimates of annual emissions of carbon (Table 2). Annual emissions increased from zero before 1960 to between 0.1 and 0.3 Pg C yr^{-1} in the 1990s (Fig. 4). The average rate of carbon release for the 1990s varied between 0.102 and $0.264 \text{ Pg C yr}^{-1}$, and the mean flux over the 10-year period (1989–98) was $0.18 \text{ Pg C yr}^{-1}$. The main contributor to flux is the decay of wood that had been chopped down. In contrast, the uptake of carbon in secondary growth, and the releases from burning associated with deforestation, are small (Fig. 4).

The average annual source of $0.18 \text{ Pg C yr}^{-1}$ from deforestation and regrowth is near the low end of the range of estimates from previous analyses of Brazil (0.174 – $0.336 \text{ Pg C yr}^{-1}$; refs 7, 27, 28), and lower than the estimate for Brazilian Amazonia ($0.261 \text{ Pg C yr}^{-1}$; ref. 22). Many of these earlier estimates were for a single year or an average year, were based on tabular rather than spatial data, used a small number of average values for biomass and did not take into account the sources and sinks embodied in previous land-use history. Fearnside's recent analysis²² is the most comprehensive, but his estimate is difficult to compare with ours because he computed a 'committed' flux that included future sources and sinks embodied in current slash pools and regrowing forests. The estimate determined here lowers the global estimate of flux for the 1980s (ref. 8) by about 0.2 Pg C yr^{-1} .

The greatest uncertainty in the calculated flux of carbon over the period 1989–1998 results from uncertainty in the biomass of forests deforested. Uncertainty in biomass, rates of deforestation and rates of decay accounted for about 60, 25 and 15%, respectively, of the range of estimates of flux. Rates of decay were relatively more important in the years following increased rates of deforestation.

The average biomass of forests was 145 Mg C ha^{-1} for the first estimate obtained from RADAMBRASIL data, and 232 Mg C ha^{-1} (60% higher) for the second estimate. The reasons for the discrepancy include different adjustments for below-ground biomass,

dead biomass, palms, vines, and understory vegetation, and the possibilities of bias in sampling²⁹. The third estimate of biomass, based on independent field measurements, yielded an intermediate value of 210 Mg C ha^{-1} . Despite the large uncertainty in estimates, average biomass of the forests actually deforested was only 4–6% less than the average biomass of the region's forests, whichever estimate was used.

The deforested area reported by INPE in 1988 was about 25% higher than the deforested area interpolated between 1986 and 1992 classifications of land cover with Landsat data (Fig. 2). Both estimates of deforestation were obtained from Landsat data, and it is not clear which estimate is more accurate. INPE uses a manual approach for delineating areas deforested, whereas our land-cover classifications were based on supervised, automated techniques using digital data.

Although logging and fire were not considered in this analysis, they are important in the region and affect the storage of carbon. The net effect of logging is probably small (4–7% of the release from deforestation)⁴, because releases of carbon associated with harvest are largely offset by the accumulations in forests recovering from earlier harvests. The area annually burned is unknown, but fire may release as much carbon as is released from deliberate deforestation, particularly following the dry years associated with an El Niño event^{4,30}.

The combined effects of deforestation, abandonment, logging and fire may thus yield sources of carbon that vary between 0.1 and 0.4 Pg C yr^{-1} . These fluxes are similar in magnitude (but opposite in sign) to the sink calculated recently for natural ecosystems in the

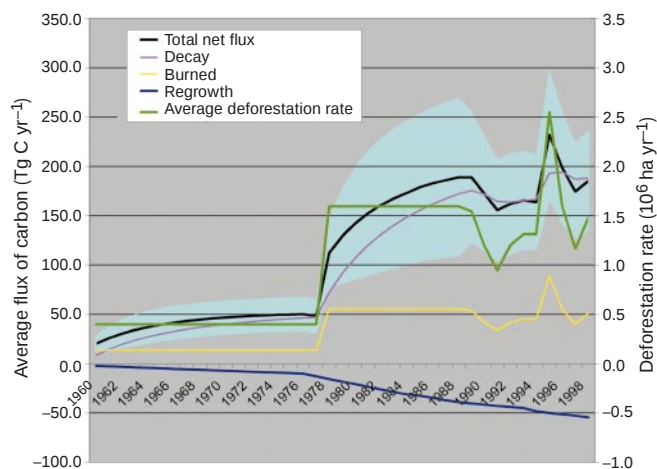


Figure 4 Annual rate of deforestation and mean annual sources and sinks of carbon in Brazilian Amazonia. Shaded area is ± 1 s.d. from the mean annual flux of carbon determined for the eight cases described in the text.

region^{5,6}. Taken together, the sources (from land-use change and fire) and the sinks (in natural forests) suggest that the net flux of carbon between Brazilian Amazonia and the atmosphere may be nearly zero, on average. However, the interannual variability of the natural fluxes, including fire, is larger than it is for the human-induced sources: the annual net flux for this significant region of the tropics may vary between a sink and a source of 0.2 Pg C yr⁻¹, occasionally more. □

Received 11 August; accepted 10 November 1999.

- Ciais, P. *et al.* Partitioning of ocean and land uptake of CO₂ as inferred by $\delta^{13}\text{C}$ measurements from the NOAA Climate Monitoring and Diagnostics Laboratory global air sampling network. *J. Geophys. Res.* **100**, 5051–5070 (1995).
- Keeling, R. F., Piper, S. C. & Heimann, M. Global and hemispheric CO₂ sinks deduced from changes in atmospheric O₂ concentration. *Nature* **381**, 218–221 (1996).
- Rayner, P. J., Enting, I. G., Francey, R. J. & Langenfelds, R. Reconstructing the recent carbon cycle from atmospheric CO₂, $\delta^{13}\text{C}$ and O₂/N₂ observations. *Tellus B* **51**, 213–232 (1999).
- Nepstad, D. C. *et al.* Large-scale impoverishment of Amazonian forests by logging and fire. *Nature* **398**, 505–508 (1999).
- Prentice, I. C. & Lloyd, J. C. Quest in the Amazon Basin. *Nature* **396**, 619–620 (1998).
- Tian, H. *et al.* Effect of interannual climate variability on carbon storage in Amazonian ecosystems. *Nature* **396**, 664–667 (1998).
- Houghton, R. A. *et al.* The flux of carbon from terrestrial ecosystems to the atmosphere in 1980 due to changes in land use: geographic distribution of the global flux. *Tellus B* **39**, 122–139 (1987).
- Houghton, R. A. The annual net flux of carbon to the atmosphere from changes in land use 1850–1990. *Tellus B* **51**, 298–313 (1999).
- Fisher, M. J. *et al.* Deep-rooted grasses store carbon in South American soils. *Nature* **371**, 236–238 (1994).
- Davidson, E. A., Nepstad, D. C., Klink, C. & Trumbore, S. E. Pasture soils as a carbon sink. *Nature* **376**, 472–473 (1995).
- Neill, C. *et al.* Forest- and pasture-derived carbon contributions to carbon stock and microbial respiration of tropical pasture soils. *Oecologia* **107**, 113–119 (1996).
- Fearnside, P. M. & Barbosa, R. I. Soil carbon changes from conversion of forest to pasture in Brazilian Amazonia. *For. Ecol. Mgmt* **108**, 147–166 (1998).
- Fearnside, P. M. *Human Carrying Capacity of the Brazilian Rainforest* (Columbia Univ. Press, New York, 1986).
- Nepstad, D. C., Uhl, C. & Serrão, E. A. S. Recuperation of a degraded Amazonian landscape: Forest recovery and agricultural restoration. *Ambio* **20**, 248–255 (1991).
- Moran, E., Brondizio, E., Mausel, P. & Wu, Y. Deforestation in Amazonia: land use change from ground and satellite level perspectives. *BioScience* **44**, 329–338 (1994).
- Skole, D. L., Chomentowski, W. H., Salas, W. A. & Nobre, A. D. Physical and human dimensions of deforestation in Amazonia. *BioScience* **44**, 314–321 (1994).
- Alves, D. S. & Skole, S. L. Characterizing land cover dynamics using multi-temporal imagery. *Int. J. Remote Sensing* **17**, 835–839 (1996).
- Brazil, Ministério das Minas e Energia, Departamento Nacional de Produção Mineral (DNPM), Projeto RADAMBRASIL: 1973–1983. *Levantamento de Recursos Naturais*, Vols. 1–23, DNPM, Rio de Janeiro.
- Brown, S. & Lugo, A. E. Aboveground biomass estimates for tropical moist forests of the Brazilian Amazon. *Interciencia* **17**, 8–18 (1992).
- Brown, S., Gillespie, A. J. R. & Lugo, A. E. Biomass estimation methods for tropical forests with applications to forest inventory data. *For. Sci.* **35**, 881–902 (1989).
- Cairns, M. A., Brown, S., Helmer, E. H. & Baumgardner, G. A. Root biomass allocation in the world's upland forests. *Oecologia* **111**, 1–11 (1997).
- Fearnside, P. M. Greenhouse gases from deforestation in Brazilian Amazonia: net committed emissions. *Clim. Change* **35**, 321–360 (1997).
- Fearnside, P. M., Leal, N. & Fernandes, F. M. Rainforest burning and the global carbon budget: Biomass, combustion efficiency, and charcoal formation in the Brazilian Amazon. *J. Geophys. Res.* **98**, 16733–16743 (1993).
- Carvalho, J. A., Higuchi, N., Araujo, T. M. & Santos, J. C. Combustion completeness in a rainforest clearing experiment in Manaus, Brazil. *J. Geophys. Res.* **103**, 13,195–13,199 (1998).
- Delaney, M., Brown, S., Lugo, A. E., Torres-Lezama, A. & Bellow Quintero, N. The quantity and turnover of dead wood in permanent forest plots in six life zones of Venezuela. *Biotropica* **30**, 2–11 (1998).
- Brown, S. & Lugo, A. E. Tropical secondary forests. *J. Trop. Ecol.* **6**, 1–32 (1990).
- Hall, C. A. S. & Uhlir, J. Refining estimates of carbon released from tropical land-use change. *Can. J. For. Res.* **21**, 118–131 (1991).
- Schroeder, P. E. & Winjum, J. K. Assessing Brazil's carbon budget: II. Biotic fluxes and net carbon balance. *For. Ecol. Mgmt* **75**, 87–99 (1995).
- Fearnside, P. M. Forest biomass in Brazilian Amazonia: Comments on the estimate by Brown and Lugo. *Interciencia* **17**, 19–27 (1992).
- Cochrane, M. A. *et al.* Positive feedbacks in the fire dynamic of closed canopy tropical forests. *Science* **284**, 1832–1835 (1999).

Acknowledgements

We thank A. Janetos, D. Nepstad, C. Prentice and R. Birdsey for comments on earlier drafts of the paper. This work was supported by the Terrestrial Ecology Program, the Land Cover and Land Use Change Program, and the Landsat Program in the NASA Office of Earth Science.

Correspondence and requests for materials should be addressed to R.A.H. (e-mail: rhoughton@whrc.org).

Rapid evolution of male reproductive genes in the descent of man

Gerald J. Wyckoff*, Wen Wang† & Chung-I Wu*†

* Committee on Genetics and † Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637, USA

A diverse body of morphological and genetic evidence has suggested that traits pertaining to male reproduction may have evolved much more rapidly than other types of character^{1–3}. Recently, DNA sequence comparisons have also shown a very high level of divergence in male reproductive proteins between closely related *Drosophila* species^{4–6}, among marine invertebrates^{7,8} and between mouse and rat⁹. Here we show that rapid evolution of male reproductive genes is observable in primates and is quite notable in the lineages to human and chimpanzee. Nevertheless, rapid evolution by itself is not necessarily an indication of positive darwinian selection; relaxation of negative selection is often equally compatible with the DNA sequence data. By taking three statistical approaches, we show that positive darwinian selection is often the driving force behind this rapid evolution. These results open up opportunities to test the hypothesis that sexual selection plays some role in the molecular evolution of higher primates.

Although positive (or directional) selection may justifiably be considered the essence of darwinian evolution, proving its action at the molecular level is often difficult, especially in humans. In the rare cases when positive selection is strong enough to rise above the background of neutral evolution or overcome other forms of selection, the rate of non-synonymous nucleotide substitution (K_A) may exceed that of synonymous substitution (K_S)¹⁰. Many of the genes cited above that function primarily in males' reproductive tissues do indeed have $K_A > K_S$. They not only suggest positive selection but also implicate selection "in relation to sex"¹¹. Evidence for sexual selection has been reported extensively in previous morphological and genetic analyses^{1–3}.

Here we attempt to find out how strong positive selection (especially in relation to sexual functions) has been in the lineage leading to modern human by analysing a cluster of three spermatid-associated protein genes and then a broad class of genes of male reproduction. We use three approaches to reveal the action of positive selection: (1) showing that the K_A/K_S ratio is significantly greater than one; (2) contrasting the pattern of intraspecific variation with that of interspecific divergence¹²; and (3) classifying amino-acid substitutions into categories based on the physicochemical properties of the residues^{13,14}. The current wealth of DNA sequence data is particularly suitable for this third approach.

Table 1 The numbers of replacement (R), silent (S) and noncoding (S') nucleotide changes

	R	S+S'	(S)
<i>Pm-1</i>			
Within (<i>n</i> = 52)	0	6	(2)
Between	9	9	(0)
No. of sites	114	483	(39)
		<i>P</i> = 0.037	(<i>P</i> = 0.018)
<i>Pm-2</i>			
Within (<i>n</i> = 32)	0	9	(0)
Between	7	11	(2)
No. of sites	224	705	(82)
		<i>P</i> = 0.035	(NC)

Within denotes 'within human populations'; *n* indicates the number of haploid genomes sampled; Between denotes 'between human and chimpanzee'. In this application of the McDonald and Kreitman test¹², we compare *R* with both *S+S'* and *S* alone (in parentheses). The *P* values given are from the one-tailed Fisher's Exact test. Total numbers of sites for each class by the method of Li *et al.*³⁰ are also given.

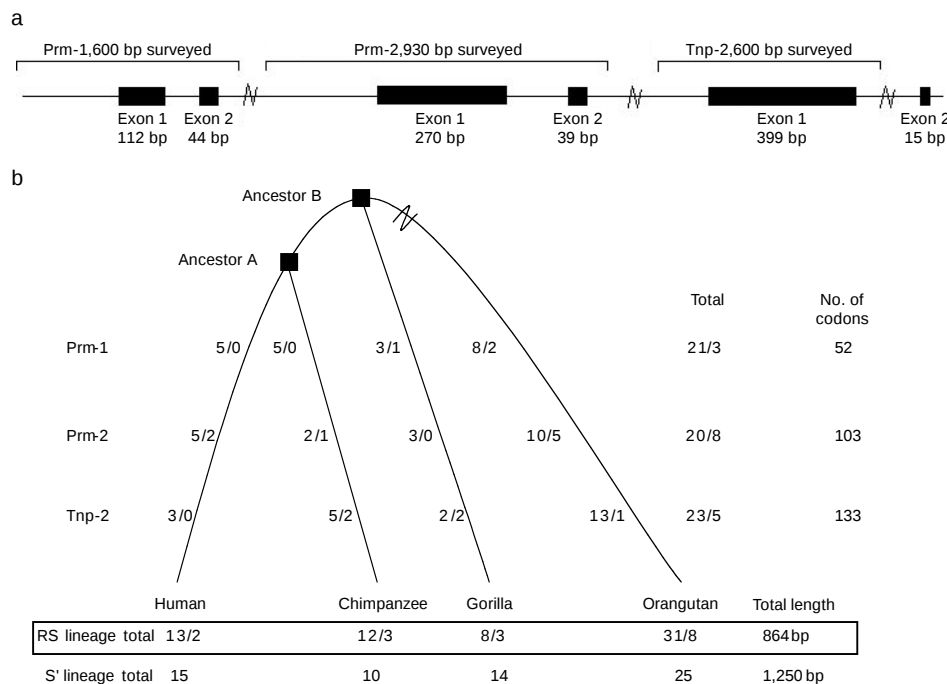


Figure 1 Evolution of the protamine gene complex. **a**, The structure of the region encoding the Prm-1, Prm-2 and Tnp-2 genes. The region spans 17 kb and the specific regions sequenced are indicated above the gene map. Sample sizes are given in Methods. **b**, The observed numbers of replacement (*R*, boldface) and synonymous (*S*) nucleotide changes for Prm-1, Prm-2, and Tnp-2 are given along each branch. *S'* is for changes in

the non-coding region. The branch to orangutan is between ancestor B and the extant species whereas that to each of the others (including gorilla) is from ancestor A. Chimpanzee here includes both chimpanzee and bonobo. Old World monkey sequences are used as the outgroup. Both the lineage and genic sums are also given.

We first studied three spermatid-associated protein genes, Protamines 1 and 2 (*Prm-1* and *Prm-2*, respectively) and Transitional protein 2 (*Tnp-2*), which are tightly clustered in a 17-kilobase (kb) region on chromosome 16p13.3 in humans¹⁵. The regions analysed are depicted in Fig. 1a. As primates' sperm morphology evolves rapidly in shape, volume and length¹⁶, the two protamines may potentially influence the sperm's morphology, and ultimately its competitiveness in fertilizing the egg. TNP-2, although necessary for spermatogenesis, is not present in the mature sperm. We re-analyse the synonymous (*S*) and amino-acid replacement (*R*) nucleotide substitutions of *Prm-1* (ref. 17), *Prm-2* (ref. 18) and *Tnp-2* (ref. 15) along each branch leading to the extant great apes (Fig. 1b; see Methods). The underlying hypothesis is that potential competition

among sperm from different males has contributed to the accelerated evolution of genes involved in sperm and seminal fluid production. Figure 1b shows two interesting trends. First, PRM-1, the smallest of the three proteins and the only one that is present in sperm in its entirety, has experienced the highest rate of non-synonymous substitution. For example, among human, chimpanzee and gorilla, the *R/S* ratio is 13/1. Second, the *R/S* ratios appear higher in the human and chimpanzee lineages than in gorilla. The contrast is intriguing in light of the social-sexual behaviours of the African apes^{19,20}. Whereas modern chimpanzees and bonobos are clearly promiscuous with ample chance of multiple insemination, ovulating female gorillas seem much less likely to be multiply inseminated. (We are less certain about orangutan

Table 2 Male reproduction-associated genes

	Name	Size (bp)	$K_A (\times 10^3)$	$K_S (\times 10^3)$	K_A/K_S
$K_A > 0.05$	Acrosin-Trypsin inhibitor	255	13.80	6.40	2.14
	Protamine 1	150	13.30	4.60	2.89
	PSP94	342	10.50	9.90	1.06
	Protamine 2	309	9.70	7.20	1.35
	Transitional Protein 2	387	5.60	9.20	0.61
	SRY	615	5.50	9.60	0.58
	Histone H1 (testicular)	620	5.50	9.60	0.58
	Prostate Specific Antigen (PSA)	786	5.40	9.60	0.56
	Ph-20	1530	5.20	9.20	0.57
	Fertilin Beta	2207	5.10	5.70	0.90
$K_A < 0.05$	TSPY	640	5.10	4.42	1.10
	SP10	747	4.00	11.00	0.36
	Steroid 5-Alpha reductase Type 1	777	3.50	6.50	0.57
	Steroid 5-alpha reductase Type 2	765	2.30	5.20	0.44
	SP17	453	1.70	1.30	1.31
	FSHR	2088	1.00	3.80	0.27
	Androgen receptor	2267	0.40	5.00	0.09
	Epi-1	456	0.00	6.00	0.00

Divergence of genes that are primarily associated with male reproduction between humans and Old-World monkeys. K_A , number of non-synonymous substitutions per site; K_S , number of synonymous substitutions per site.

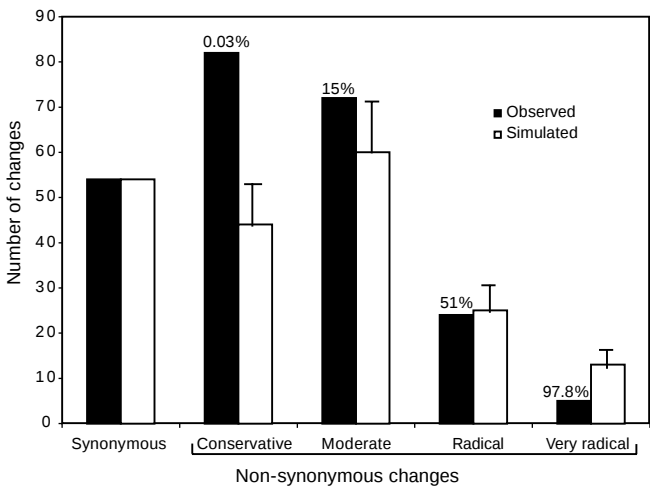


Figure 2 Observed and simulated numbers of nucleotide changes. Data are partitioned into five classes, for the genes of Tables 2 and 3 whose K_A/K_S exceed 1. The simulated synonymous changes are set equal to the observed. Non-synonymous changes are further partitioned into four classes according to the physico-chemical differences of the underlying amino acids¹³. The percentage proportion of simulated cases that have a higher number of changes than the observed cases is given. The error bar represents 1 s.d.

because the longer branch to this species makes the inference of substitutions less accurate and also because the mating behaviour of modern orangutan is not precisely defined. Nevertheless, the orangutan testis/body weight ratio is closer to the human ratio than to the ratio for gorilla (p. 219 of ref. 19.) We shall conduct statistical tests with additional data below.

That protamines evolve rapidly has been previously demonstrated^{17,18,21}. The question is whether positive selection needs to be invoked to account for the high rate of amino-acid substitution among apes (or, in other words, whether the alternative

explanations of relaxation of selective constraints and elevation of mutation rate are insufficient²¹). Some earlier studies have accepted the former alternative^{17,18}. We sequenced the contiguous untranslated region or introns shown in Fig. 1a and designated changes in these regions S' . As the divergence estimated from both S and S' is about 1.6% between human and chimpanzee, the mutation rate in this gene cluster is almost identical with the genomic average^{22,23}. This point is relevant because, for a small gene like *Prm-1*, K_S can be smaller than average by chance and $K_A > K_S$ would not necessarily indicate positive selection. It can then be shown, by computer simulation (see Methods), that the observed numbers of non-synonymous changes in the three genes combined are significantly higher than expected in the human lineage and between human and chimpanzee ($P < 0.005$ in both cases). In contrast, the sums of synonymous and non-coding substitutions are nearly as expected. Moreover, when all changes that may potentially be influenced by methylation and deamination in CpG dinucleotides are removed from consideration, the level of significance remains at $P < 0.01$ for the human–chimpanzee comparison and $P < 0.05$ for the human lineage. This factor has often been neglected²¹.

A more sensitive test is based on the expectation that, if relaxation of negative selection is responsible for the rapid amino-acid replacements between species, the level of replacement polymorphism within species should increase accordingly. This is not expected under positive (directional) selection because substitutions would then have a short retention time in the population and hence contribute little to the polymorphism. The McDonald and Kreitman (MK) test¹² was developed to distinguish between these possibilities by contrasting the levels of R and S within and between species in a 2×2 contingency table. If positive selection drives the bulk of amino acid substitutions, then the R/S ratio should be significantly higher for the between-species divergence than for the within-species polymorphism. In the calculation, we also include S' (silent changes in the contiguous introns or untranscribed regions) with S . The test demands that nucleotide sites for R and S (or S') do not consistently have different histories. As each gene sequence is small (600–930 base pairs) and no intragenic recombination is detected, the assumption appears to be valid.

Table 3 Reference group genes

	Name	Size (bp)	$K_A (\times 10^3)$	$K_S (\times 10^3)$	K_A/K_S
$K_A > 0.05$	Glycophorin A	423	15.10	6.40	2.36
	Lysozyme	444	6.50	2.30	2.83
	RH-50	1227	6.10	4.60	1.30
	Leptin	501	5.50	6.60	0.83
$K_A < 0.05$	CD4	1192	4.60	6.60	0.64
	Kallikrein	789	3.90	5.20	0.73
	Carboxylesterase	1701	3.60	8.00	0.45
	Prepromotilin	345	3.40	1.70	2.04
	ZNF80	816	3.40	3.20	1.07
	Interferon Gamma	399	3.20	5.40	0.59
	ZP3	1275	3.00	7.60	0.40
	Interleukin 16	1890	2.50	6.10	0.40
	CA1	783	2.50	7.70	0.33
	Oviductal Glycoprotein	1824	2.30	5.30	0.44
	P53	1182	2.20	7.70	0.28
	K6	453	2.10	5.40	0.38
	Cholesteryl ester transferase	1482	2.00	7.00	0.29
	Beta-globin	429	1.80	6.00	0.30
	Interleukin 6	639	1.80	4.80	0.38
	Oxytocin receptor	1167	1.50	8.10	0.18
	GPR1	1065	1.30	4.40	0.31
	Cystatin C	438	1.30	11.40	0.11
	Galactocerebrosidase	2007	1.30	5.90	0.22
	Epsilon-globin	444	1.30	2.00	0.67
	CCR5	1060	1.10	4.50	0.25
	CFTR	4446	0.80	6.20	0.12
	CXCR4	1060	0.80	6.20	0.12
	Dopamine D1	1341	0.10	7.80	0.01
	IRK 1	1221	0.00	5.00	0.00
	Cyclophilin A	495	0.00	3.90	0.00

Divergence of 30 genes selected without regard to function from Genebank between humans and Old-World monkeys. K_A , number of non-synonymous substitutions per site; K_S , number of synonymous substitutions per site.

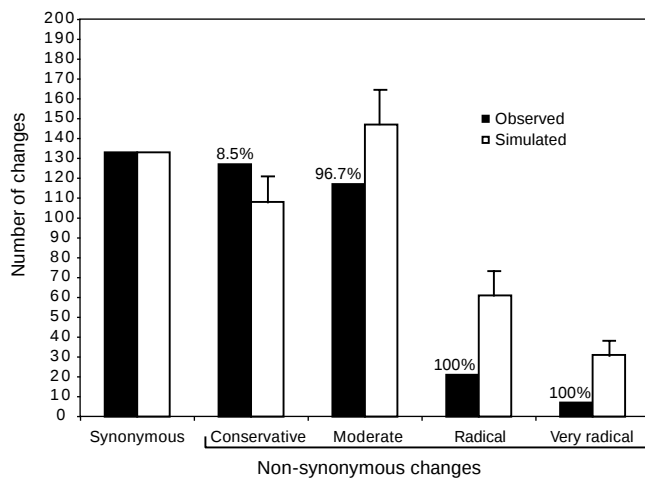


Figure 3 Observed and simulated numbers of nucleotide changes. Data are partitioned into five classes, for the faster evolving 11 male reproductive genes of Table 2. See also Fig. 2.

We compare human and chimpanzee to determine the divergence of the protamine genes and survey human populations for polymorphism. Table 1 shows a significant deficit in *R* within human populations for both *Prm-1* and *Prm-2*. For *Prm-1*, the deficit may in fact be more significant than shown here. In two previous surveys of the exon sequences from a total of 78 individuals^{24,25}, no replacement polymorphism in human populations was detected. As complete sequences from the intron or non-coding region are unavailable from these individuals, the within-species *R/(S+S')* could be more biased than the 0/6 ratio of Table 1 if these sequences were to be included. Overall, the excess of replacements between human and chimpanzee in both protamines corroborates the hypothesis that positive darwinian selection is responsible for the high rate of amino-acid replacements. (*Tnp-2*, with only three polymorphic sites in our survey, is uninformative in the MK test.)

Although the MK test is clearly a more discerning test than the interspecific comparison of *R* and *S* alone in detecting positive selection, the need to measure within-species nucleotide polymorphism severely limits the number of genes that can be examined. In addition, if a survey does not find enough variable sites (as in our survey of *Tnp-2*), the MK test is uninformative. Also, the level of polymorphism is in fact expected to be low after a recent round of positive selection, a process referred to as selective sweep. In what follows, we extend an earlier method¹³ that could enhance the power of detecting positive selection but uses only interspecific comparisons.

The estimation of *R/S* places all non-synonymous nucleotide substitutions into a single category even though they are likely to be highly heterogeneous in their fitness consequences. Li *et al.*¹³ classified non-synonymous substitutions into categories according to the similarity of amino acids, as measured by Grantham's distance²⁶. Grantham's distance relies on size, polarity, and carbon composition of the side-chains of amino acids to measure the underlying physico-chemical similarity between residues. Substitutions with large Grantham's distances have been shown to be much less frequent than expected¹³, a result attributed to greater negative selection against radical changes. We adopt Li *et al.*'s practice of classifying the observed codon substitutions as synonymous, conservative (≤ 50), moderate (51–100), radical (101–150) or very radical (>150) amino-acid changes. The numbers in parentheses denote the range of Grantham's distances within each category. In Methods, we describe the procedure of computing the expected numbers for the five classes of substitution (from synonymous to very radical amino-acid changes) as well as simulating the distribu-

tion under the strict assumption of no selection against any substitution. The expected number of synonymous substitutions in the simulations is set to equal the observed.

The objective of this test is to detect positive selection in the data of Tables 2 and 3. We posit that the signature of positive selection is more likely to be detected among conservative amino-acid substitutions for two reasons. First, there would be less negative selection on these substitutions according to the neutral theory¹⁰. Second, there may also be more positive selection on these substitutions according to the neo-darwinian view²⁷ in which selection generally favours small changes over big leaps. Slowly evolving genes like somatic histones, however, should be excluded from the test because they have probably experienced strong negative selection in all classes of amino-acid change. We thus examined the 11 genes in Tables 2 and 3 that have a K_A/K_S value greater than 1, most of them insignificant. The null hypothesis by the neutral theory predicts that the observed/expected ratios for the four classes would not be significantly different from 1. The observation versus prediction as shown in Fig. 2 rejects this null hypothesis. The ratios, from the synonymous to the highly radical class, are: 1, 1.86, 1.20, 0.96 and 0.39. The excess in the conservative class is significant with $P = 0.0003$ whereas the deficit in the highly radical class is significant at $P = 0.022$. It seems clear that, among rapidly evolving genes, the forces of positive and negative selection operate simultaneously but their relative strength diminishes as the functional effect of substitution increases. This conclusion is supported by surveys of protein evolution that take into account physico-chemical properties of amino-acid changes^{13,14}. Although the general pattern may often hold true when the procedure is applied to specific genes (as in the case of *Drosophila* ACP26Aa⁴, the method is meant to portray the statistical average over many genes.

To estimate the extent of positive selection driving the evolution of male reproductive genes, we examined the faster evolving 50% of the genes of Table 2, weighted by their length. In doing so, we included the first 11 genes, which account for 80% of the codon changes observed between human and Old World monkey (OWM) in Table 2. The lowest K_A for the 11 genes is 0.05, about 70% of the genome-wide average of 0.07 substitutions per base pair between human and OWM^{22,23}. In comparison, only 4 of the 30 genes have a K_A value of 0.05 or greater in the reference group of Table 3, which represents a crude 'genomic average'. In general, there appear to be more genes in Table 2 that can be considered moderately or rapidly evolving than the genomic average, although a statistical test is not meaningful owing to the possibly non-random representation of genes in the data bank.

For the 11 rapidly evolving genes of Table 2, the observed (expected) numbers for each of the five classes are 133 (133), 127 (108), 117 (147), 21 (61) and 7 (31), respectively, as shown in Fig. 3. The respective ratios are 1, 1.18, 0.80, 0.34 and 0.23. The observed number of conservative amino-acid substitutions is 18% higher than the expected whereas the other classes suffer reductions that are collectively far greater than the gain. As a result, the overall K_A is lower than K_S for this set of genes. If each class is considered separately, the deficit in amino-acid changes is significant for all three non-conservative classes but the excess for the conservative class is only marginally so with $P = 0.085$. However, this probability is computed on the assumption that there is no negative selection at all between conservative amino-acid changes. The assumption seems overly stringent considering the observed pattern of the reference group of genes (see below). Moreover, Fig. 3 includes many genes that are under substantial selective constraints.

We further show that the pattern of Figs 2 and 3 (that is, conservative amino-acid substitutions are more frequent than expected) is neither an average pattern of the genome nor an artefact of the analysis. For the 30 genes of Table 3, the observed (expected) numbers for each of the five classes are 461 (461), 186 (369), 177 (494), 43 (195) and 14 (95). The observed/expected

ratios are 1, 0.50, 0.36, 0.22 and 0.15. This 'average pattern' shows substantial selective constraint even between conservative amino-acid changes. It is certainly not true that positive darwinian selection can only be observed among male reproductive genes. Many genes of non-reproductive functions must also be subjected to it. Unfortunately, these genes are statistically more difficult to study because of the heterogeneity in their functions and evolutionary dynamics. In this regard, we notice that the three fastest-evolving genes in Table 3 are all at the front line of interaction with foreign substances. Future studies of different classes of genes (for example, cell-surface antigens, receptors or seminal-fluid proteins) may provide further insight.

Among genes of male reproduction, protamines are unique in that they have functional analogues in somatic histones. Whereas the latter are among the slowest evolving genes in the human genome, protamines are at the extreme opposite end of the spectrum. Another example of such a contrast from *Drosophila* has been reported⁵. In the protamine gene cluster, there also appears to be a correlation between the molecular evolutionary pattern and the mating system in extant species. Whether this is a general trend will require more extensive comparisons between genes of promiscuous species, such as bonobos, and their non-promiscuous relatives. If the putative correlation is general, molecular data might shed some light on the controversial issues of the socio-sexual structure of early humans and ancestral primates^{19,20}.

We also wish to distinguish this current hypothesis of selection-driven evolution of male reproductive genes from the phenomenon of male mutation-driven evolution²⁸. In the latter, genes that pass through males would have a higher mutation rate. The hypothesis addressed in this report, on the other hand, is about selection and function and is independent of the origin of mutations.

It is significant that the faster-evolving 11 of the 18 male reproductive genes collectively bear a detectable signature of Darwinian positive selection between man and OWM. Because these genes account for 80% of the total observed changes within our pool of data, the result suggests pervasive positive selection with regard to male reproduction. That only the conservative class of amino-acid substitutions experience accelerated evolution is compatible with the neo-darwinian view of incremental functional improvements during evolution²⁷. □

Methods

DNA sequences

The coding region sequences of *Prm-1*, *Prm-2* and *Tnp-2* of human, apes and monkeys are from the literature^{15,17,18}. We also sequenced 600 base pairs of *Prm-1* from 26 humans, 4 chimpanzees, one bonobo and 2 gorillas and 930 base pairs of *Prm-2* from 16 humans and one chimpanzee, one bonobo and one gorilla. Both sets of sequences encompass the coding regions (156 base pairs and 309 base pairs, respectively) and the adjacent noncoding or intron regions (Fig. 1a). In addition, we sequenced 600 base pairs from the *Tnp-2* gene from 10 humans, one chimpanzee, one bonobo and one gorilla. For the human polymorphism survey, African-Americans constitute about 80% of the samples with Caucasians and Asians accounting equally for the remaining 20%. DNA samples for the study were subjected to polymerase chain reactions using several sets of primers (details upon request). Sequencing reactions were carried out at least twice for each strand on ABI 377.

When gathering genes for Tables 2 and 3, our primary concerns were completeness of the sequence and availability of information regarding the gene's function and expression. Genes that are expressed primarily in male reproductive tissues were grouped in Table 2. Others were grouped into Table 3. Genes that contain trans-specific polymorphism, such as the major histocompatibility complex region or immunoglobulins, are not included in this report.

Analysis and simulations

The assignment of non-synonymous and synonymous changes to the branches of Fig. 1b was done by using the old-world monkey sequences as the outgroup. In the event of ambiguity changes were assigned so as to minimize the number among human, chimpanzee and gorilla. Because of the small sizes of *Prm-1*, *Prm-2* and *Tnp-2* as well as their unusual amino-acid compositions, the standard K_A/K_S analysis does not present a suitable statistical test. Instead, we carried out computer simulations as follows. (1) For each of the one million replicates simulated, the number of nucleotide changes is drawn from a Poisson distribution with its mean equal to the published estimates of divergence

for these branches. For the human–chimpanzee lineage, the mean divergence was 1.6%. (2) Using the ancestral sequence between humans and chimpanzees as a starting point, the position of change in the sequence was randomly chosen. The change was then assigned as a transition or transversion according to published estimates in mammals where 60–67% are transitions^{22,23,29}. We use 60% in our simulations. For the coding regions, changes were then determined to be either R or S. Stop codon changes are not counted. This step is sequentially repeated for the number of mutations obtained in step 1. The results were tallied for each replicate and the distributions were used to determine the *P*-value for the observed R, S and S'.

In Tables 2 and 3, K_A and K_S values were obtained by using the method of ref. 30. In Figs 2 and 3, we classified all one-base pair codon changes between human and OWM genes into five classes¹³ (synonymous and four classes of non-synonymous changes; see text). Only changes that occurred in single-hit codons were used in this analysis. The numbers of changes and their distributions under strict neutrality are obtained for all classes as follows. All sequences from human and OWM are concatenated and changes are introduced one by one. Each change is first assigned randomly to a position and then determined to be either a transition or transversion. (Transitions are set to be 60% of the total changes^{22,23,29}, which is slightly conservative in this application.) Each codon, in one step, can change in nine different ways, which may include 0–3 synonymous changes and 6–9 non-synonymous changes. Each non-synonymous change is further classified into one of four bins by the physico-chemical differences of the coded amino acids^{13,26}. The resulting codon change is substituted into the starting sequence (unless it is a stop codon) and the type of change is recorded into one of the five bins. The process is repeated until the total number of synonymous changes equals the observed. Ten thousand replicates were done and the mean and standard deviation for each of the four non-synonymous classes are presented.

Received 17 February; accepted 5 November 1999.

- Eberhard, W. G. *Sexual Selection and Animal Genitalia* (Harvard Univ. Press, Cambridge, MA, 1985).
- Coulthart, M. B. & Singh, R. S. Low genic variation in male reproductive tract proteins of *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **5**, 167–181 (1988).
- Wu, C. I., Johnson, N. A. & Palopoli, M. F. Haldane's rule and its legacy: why are there so many sterile males. *Trends Ecol. Evol.* **11**, 281–284 (1996).
- Tsaur, S., Ting, C. T. & Wu, C. I. Positive selection driving the evolution of a gene of male reproduction, ACP26Aa of *Drosophila*: II Divergence vs. polymorphism. *Mol. Biol. Evol.* **15**, 1040–1046 (1998).
- Ting, C., Tsaur, S. C., Wu, M. L. & Wu, C. I. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282**, 1501–1504 (1998).
- Nurminsky, D. I., Nurminskaya, M. V., Aguiar, D. D. & Hartl, D. L. Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. *Nature* **396**, 572–575 (1998).
- Lee, Y., Ohta, T. & Vacquier, V. D. Positive selection is a general phenomena in the evolution of abalone sperm lysin. *Mol. Biol. Evol.* **6**, 424–435 (1995).
- Metz, E. C. & Palumbi, S. R. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Mol. Biol. Evol.* **13**, 397–406 (1996).
- Sutton, K. A. & Wilkinson, M. F. Rapid evolution of a homeodomain: evidence for positive selection. *J. Mol. Evol.* **45**, 579–588 (1997).
- Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, 1983).
- Darwin, C. *The Descent of Man and Selection in Relation to Sex* (D. Appleton, New York, 1871).
- McDonald, J. & Kreitman, M. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
- Li, W. -H., Wu, C. I. & Luo, C. C. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **2**, 150–174 (1985).
- Yang, Z., Nielson, R. & Hasegawa, M. Models of amino acid substitution and applications to mitochondrial protein evolution. *Mol. Biol. Evol.* **15**, 1600–1611 (1998).
- Nelson, J. E. & Krawetz, S. A. Linkage of human spermatid-specific basic nuclear protein genes. *J. Biol. Chem.* **268**, 2932–2936 (1993).
- Gould, K. G. Scanning electron microscopy of the primate sperm. *Intl Rev. Cytol.* **63**, 323–255 (1980).
- Retief, J. D. et al. Evolution of protamine P1 gene in primates. *J. Mol. Evol.* **37**, 426–434 (1993).
- Retief, J. D. & Dixon, G. H. Evolution of pro-protamine P2 genes in primates. *Eur. J. Biochem.* **214**, 609–615 (1993).
- Dixon, A. F. *Primate Sexuality* (Oxford Univ. Press, 1998).
- Sillen-Tullberg, B. & Moller, A. P. The relationship between concealed ovulation and mating systems in anthropoid primates: a phylogenetic analysis. *Am. Nat.* **141**, 1–25 (1993).
- Rooney, A. P. & Zhang, J. Rapid evolution of a primate sperm protein: relaxation of functional constraint or positive Darwinian selection. *Mol. Biol. Evol.* **16**, 706–710 (1999).
- Li, W. -H. *Molecular Evolution* (Sinauer, Sunderland, MA, 1997).
- Maeda, N., Wu, C. I., Bliska, J. & Reneke, J. Molecular evolution of intergenic DNA in higher primates: patterns of DNA changes, molecular clock and evolution of repetitive sequences. *Mol. Biol. Evol.* **5**, 1–20 (1988).
- Schlicker, M., Schnulle, V., Schnepfel, L., Vorob, V. I. & W. Engel, W. Disturbances of nuclear condensation in human spermatozoa: search for mutations in the genes of protamine 1, protamine 2 and transitional protein 1. *Hum. Reprod.* **9**, 2313–2317 (1994).
- Queral, R. et al. Direct sequencing of the human protamine P1 gene and application in forensic medicine. *J. Forensic Sci.* **38**, 1491–1501 (1993).
- Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* **185**, 862–864 (1974).
- Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
- Shimmin, L. C., Chang, B. H. -J. & Li, W. -H. Male-driven evolution of DNA sequences. *Nature* **362**, 745–747 (1993).
- Li, W. -H., Wu, C. I. & Luo, C. C. Nonrandomness of point mutations as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *J. Mol. Evol.* **21**, 58–71 (1984).
- Li, W. -H. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.* **36**, 96–99 (1993).

Acknowledgements

We thank C. Ober, C. Grimsley, C. Toomajian and P. Parham for generously providing DNA samples. We are very grateful to I. Boussy, J. Fay, M. Jensen, D. Ledbetter, W.-H. Li, M. Long and C.-T. Ting for comments on earlier drafts and to S.-C. Tsaur, J. Gladstone and M.-L. Wu for technical assistance and advice. This work was supported by NIH and NSF grants to C.I.W.

Correspondence and requests for materials should be addressed to C.I.W. (e-mail: ciwu@uchicago.edu) or G.J.W. (e-mail: gwyckoff@midway.uchicago.edu). DNA sequences of this study are under Genbank accession numbers AF215707–AF215724.

Voice-selective areas in human auditory cortex

Pascal Belin*, Robert J. Zatorre*, Philippe Lafaille*, Pierre Ahad† & Bruce Pike‡

* Neuropsychology/Cognitive Neuroscience Unit, Montreal Neurological Institute, † Psychology Department and ‡ McConnell Brain Imaging Center, McGill University, Montréal, Québec, Canada H3A 2B4

The human voice contains in its acoustic structure a wealth of information on the speaker's identity and emotional state which we perceive with remarkable ease and accuracy^{1–3}. Although the perception of speaker-related features of voice plays a major role in human communication, little is known about its neural basis^{4–7}. Here we show, using functional magnetic resonance imaging in human volunteers, that voice-selective regions can be found bilaterally along the upper bank of the superior temporal sulcus (STS). These regions showed greater neuronal activity when subjects listened passively to vocal sounds, whether speech or non-speech, than to non-vocal environmental sounds. Central STS regions also displayed a high degree of selectivity by responding significantly more to vocal sounds than to matched control stimuli, including scrambled voices and amplitude-modulated noise. Moreover, their response to stimuli degraded by frequency filtering paralleled the subjects' behavioural performance in voice-perception tasks that used these stimuli. The voice-selective areas in the STS may represent the counterpart of the face-selective areas in human visual cortex^{8,9}; their existence sheds new light on the functional architecture of the human auditory cortex.

Experiment 1 sought to identify brain regions showing higher neuronal activity during auditory stimulation with voices than with non-vocal sounds, using a functional magnetic resonance imaging (fMRI) paradigm adapted for auditory presentation (see Methods). Eight right-handed adults were scanned during silence and while they listened passively to stimuli from two categories (Fig. 1a): (1) vocal sounds produced by several speakers of different gender and age, either speech (for example, isolated words, connected speech in several languages) or non-speech (such as laughs, sighs and coughs); and (2) energy-matched, non-vocal sounds (for example, natural sounds, animal cries, mechanical sounds) from a variety of environmental sources. In all eight subjects, vocal sounds elicited significantly ($t > 5.7$, $P < 0.001$, t -test) greater activation than non-vocal sounds bilaterally in several regions of non-primary auditory cortex. The maximum of voice-sensitive activation was located along the upper bank of the central part of the STS in seven out of the eight subjects (Fig. 1b). Averaged in the group of subjects, voice-sensitive activity appeared stronger in the right hemisphere, and was distributed in three bilateral clusters along the STS (Fig. 1c): one in the anterior portion close to the temporal pole; one in the central portion approximately at the level of the anterior extension of Heschl's gyrus; and one in the STS portion posterior to Heschl's gyrus, extending dorsally and caudally to the planum temporale in

the superior temporal gyrus (Table 1). Importantly, there were no regions of significantly greater activation for non-vocal than for vocal stimuli ($t < 4.9$, $P > 0.05$).

Experiment 1 thus demonstrated that the brain contains several regions that are sensitive to voices; yet there is no clear evidence that they are selectively activated by voices. Vocal and non-vocal stimuli in expt 1 differed in a number of low-level acoustic features, and the areas activated might have simply been responding to some acoustic component more strongly present in the vocal stimuli. We therefore performed a second experiment in the same group of subjects to determine whether we would find a selective activation pattern when comparing vocal stimuli with more carefully matched control sounds. In expt 2, vocal stimuli were contrasted to four classes of control stimuli: (1) recordings of various bells, to assess whether the same regions would be activated by presentation of exemplars of a single category, rather than voices⁸; (2) human non-vocal sounds (for example, finger snaps, handclaps), to examine the possibility that the above areas would respond to any sound of human origin; (3) white noise modulated with the same amplitude envelope as the vocal sounds; and (4) scrambled voices, which preserve the amplitude envelope of vocal sounds, but sound nothing like voices (see

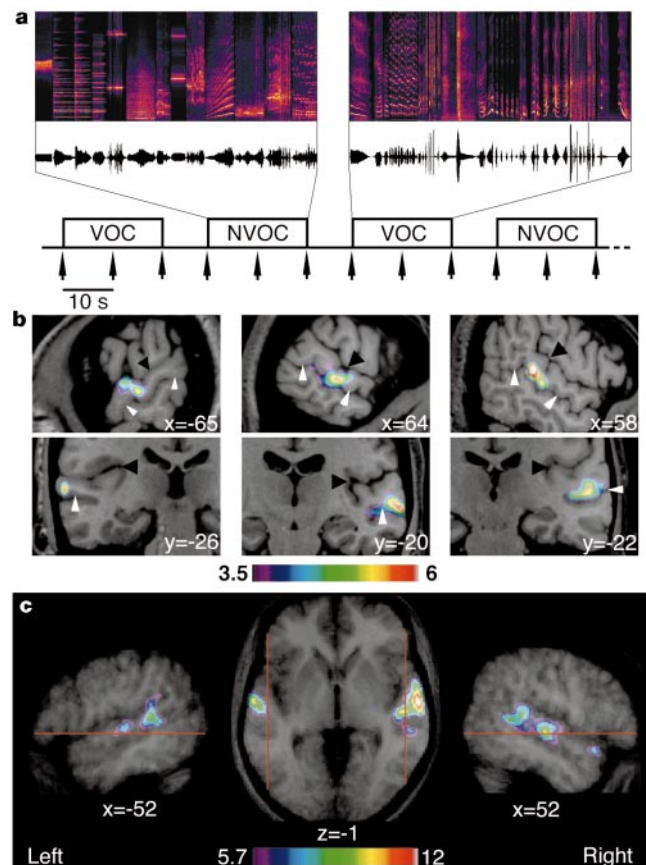


Figure 1 Experiment 1. **a**, Experimental paradigm. Spectrograms (0–4 kHz) and amplitude waveforms of examples of auditory stimuli. Vocal (VOC) and non-vocal (NVOC) stimuli are presented in 20-s blocks with 10-s silent interblock intervals, while scanning (arrows) occurs at regular 10-s intervals. **b**, Individual voice-sensitive activations. Maxima of vocal and non-vocal activation maps from three subjects are indicated in colour scale (t -value) on anatomical images in sagittal (upper panel) and coronal (lower panel) orientations (x and y , Talairach coordinates). Arrows indicate relative positions of the Sylvian fissure (black arrows) and of the superior temporal sulcus (white arrows). **c**, Voice-sensitive activation in the group average. Regions with significantly ($P < 0.001$) higher response to human voices than to energy-matched non-vocal stimuli are shown in colour scale (t -value) on an axial slice of the group-average MR image (centre, Talairach coordinate $z = -1$) and on sagittal slices of each hemisphere (sides; Talairach coordinates $x = -52$ and $x = 52$). See Table 1 and Methods for details.

Methods and Fig. 2a–c). In addition, we separated vocal speech from vocal non-speech sounds. As expected, regions of significantly greater ($t > 5.7$; $P < 0.001$) response to the vocal than to the control stimuli were again distributed along the central STS and with maxima (right: $x = 62$, $y = -14$, $z = 0$; left: $x = -58$, $y = -18$, $z = -4$; see Table 1 for description of coordinates) located close (2–7 mm) to those from expt 1. Analysis of variance shows that neuronal response in these two nearly symmetrical locations was significantly affected by stimulus class (right: $F(4,28) = 6.63$, $P < 0.001$; left: $F(4,28) = 6.49$, $P < 0.001$; Fig. 2). Interestingly, response to both speech and non-speech vocal stimuli was significantly greater than to the pooled control stimuli (right: speech $F(1,28) = 31$, $P < 0.001$; non-speech $F(1,28) = 6.8$, $P < 0.02$; left: speech $F(1,28) = 24.6$, $P < 0.001$; non-speech $F(1,28) = 4.1$, $P = 0.05$), indicating that the voice-sensitive response was not entirely due to the presence of speech in the vocal stimuli. Experiment 2 also shows that the voice-sensitive areas do not simply respond to the presentation of various sound exemplars of a same category such as bells, or to sounds of human origin, since response to voices was stronger than to these two categories, particularly in the right hemisphere (Fig. 2d, e). Moreover, expt 2 suggests that frequency structure plays a more prominent role in voice-sensitive activation than does amplitude envelope: both the amplitude-modulated noise and the scrambled voice stimuli, which elicited significantly lower responses than the vocal stimuli (left, $P < 0.015$; right, $P < 0.001$), were similar to voices in amplitude envelope but not in spectral structure (Fig. 2a–c).

A third experiment was conducted in a different group of subjects with two goals: first, to control the spectral distribution of the vocal and non-vocal stimuli, to verify that the results of expts 1 and 2 were not attributable to any difference in this variable; second, to examine how activity of the voice-sensitive areas, as well as the subjects' performance on voice-perception tasks, would be affected by modifying the spectral structure of the stimuli. Sets of vocal and non-vocal sounds were presented during scanning, both in their original form and after spectral filtering that removed either high or low frequencies, and equating vocal and non-vocal stimuli in spectral distribution (Fig. 3a). After scanning, subjects performed a vocal/non-vocal decision task and a speaker's gender-identification task that used the same filtered and unfiltered stimuli. The images corresponding to the unfiltered vocal and non-vocal sounds, averaged in the group, were very similar to those from expts 1 and 2, with voice-sensitive activity concentrated along the STS (Table 1). Maxima of voice sensitivity were located at approximately 1 cm behind and above those of expts 1 and 2 (right: $x = 54$, $y = -13$, $z = 4$; left: $x = -60$, $y = -23$, $z = 6$). Figure 3 shows that the mean activity in these regions was always greater for vocal

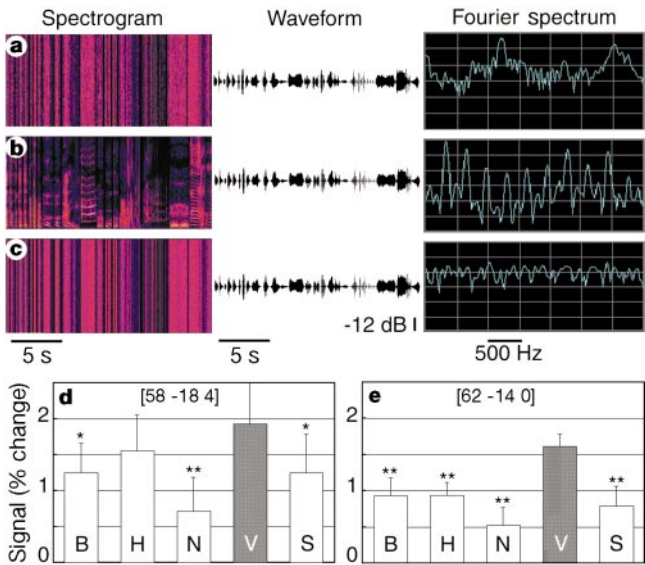


Figure 2 Experiment 2. **a–c**, Control stimuli. Scrambled vocal sounds (**a**) and amplitude-modulated white noise (**c**) have the same amplitude waveforms as the vocal stimuli (**b**), but not its characteristic spectral shape. **d, e**, Activation at voice-sensitivity maxima in expt 2. Bars (mean $\pm$ s.e.) indicate signal change for the five classes of stimuli at peak of voice selectivity in the left (**d**) and right (**e**) hemispheres. B, bells; H, human non-vocal sounds; N, amplitude-modulated noise; V, vocal sounds; S, scrambled voices; numbers in brackets: coordinates in standard stereotaxic space²⁵. Difference with vocal stimuli: asterisk, $P < 0.05$; two asterisks, $P < 0.01$. Signal changes (in per cent) are measured from the signal mean during silence.

than for non-vocal stimuli ($F(1,7) = 27.8$, $P < 0.001$), and was significantly decreased by filtering ($P > 0.05$), with no hemisphere difference ($F(1,7) = 1.46$, $P > 0.25$). Interestingly, the signal decrease was paralleled by subjects' performances on the two voice-perception tasks (Fig. 3b). Experiment 3 thus replicates the findings of expts 1 and 2, and shows that the identified regions are indeed sensitive to voice: their response to the vocal sounds was greater than to the non-vocal sounds even after their frequency distribution had been equated by filtering (Fig. 3a). Moreover, expt 3 suggests that the voice-sensitive areas might respond selectively to combination of both the high- and low-frequency components characteristic of voices¹⁰, since excluding one or the other by filtering significantly reduced both neuronal activity and subjects' accuracy on the voice-perception tasks (Fig. 3b).

In all three experiments, peaks of voice selectivity could be found in most subjects along the upper bank of the STS, a deep, long sulcus (>8 cm) running along the whole temporal lobe that is also found in many non-human primates. Activations along the STS are often reported in neuroimaging studies of human speech-processing^{11,12}, but their exact functional role remains unclear¹³. Anatomical studies in the macaque brain have shown that the STS is composed of several distinct uni- or multimodal areas, in particular, one homogeneous region lying entirely in the upper bank of the STS (area TAA)¹⁴, in a location homologous to those identified in the present study, receives its input exclusively from the auditory-related areas of the superior temporal gyrus. This area forms part of a hierarchically organized system extending anteriorly and ventrally beyond the lateral belt of the auditory cortex¹⁵, in a cortical stream that may be specialized for extracting auditory object features^{16–18}. This region is well situated to be involved in high-level analysis of complex acoustic information, such as extraction of speaker-related cues, and transmission of this information to other areas for multimodal integration and long-term memory storage¹⁹. Here, voice-selective regions were found in the STS on both sides, but voice selectivity was stronger in the right hemisphere in the first two experiments, in good agreement with the available clinical and

Table 1 Voice-sensitivity maxima

Anatomical location	Talairach coordinates			t-value		
	x	y	z	Expt 1	Expt 2	Expt 3
Right hemisphere						
STS, anterior	58	6	-10	9.2		
STS, anterior	60	-1	-4	9.6	7.9	
STS, middle	63	-13	-1	12.0	11.0	6.7
Middle temporal gyrus	52	-19	-1	9.9	7.8	9.6
STS, posterior	56	-30	6	8.0	7.2	9.7
STS, posterior	46	-44	6	9.2		
Precuneus	4	-52	30	6.0		
Left hemisphere						
Middle temporal gyrus, anterior	-60	-2	-9	6.3		
STS, middle	-62	-14	0	10.1	7.9	8.7
Planum temporale	-40	-37	13	7.1	5.9	7.2
STS, posterior	-62	-40	10	9.7		10.5

Coordinates in standard stereotaxic space²⁵ (mm) and approximate anatomical location are given for voice-sensitivity maxima of expt 1, as well as corresponding t-values above 5.7 ($P < 0.001$), for expts 1, 2 and 3. STS, superior temporal sulcus; x, lateral distances to anterior commissure–posterior commissure (AC–PC) line (positive = right); y, anterior–posterior distance to the anterior commissure (positive = in front of AC); z, distance above–below AC–PC line (positive = above).

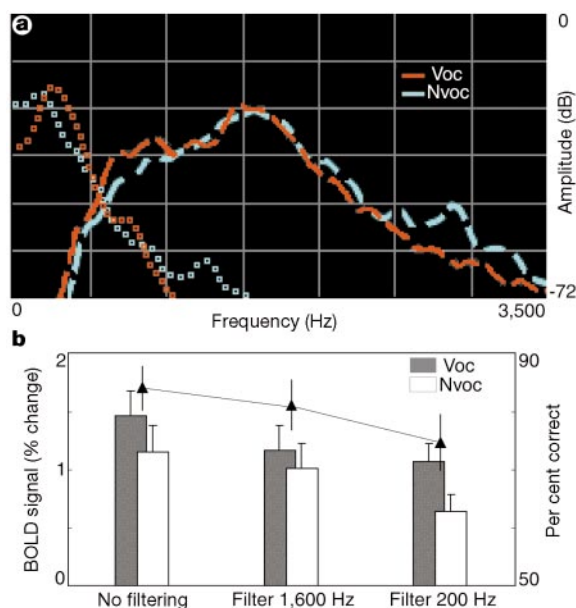


Figure 3 Experiment 3. **a**, Frequency distribution of frequency-filtered stimuli. Dashed lines, bandpass filter centred at 1,600 Hz (400 Hz bandwidth). Open squares, bandpass filter centred at 200 Hz (50 Hz bandwidth). **b**, Activation at voice-sensitivity maxima in expt 3. Bars: signal change (mean $\pm$ s.e.) for vocal stimuli is greater than for non-vocal stimuli even after their frequency distribution has been nearly equated by filtering. Behavioural performance: triangles show combined performance (mean $\pm$ s.e. in per cent correct) on the vocal/non-vocal decision task and the speaker's gender-identification task, showing a decrease with filtering that parallels that of signal change in that area.

psychological literature^{4–7,20}. However, the same pattern was not found in expt 3, suggesting that the neural substrate of voice perception might be less clearly lateralized than in the case of speech perception.

In conclusion, these experiments provide strong evidence that the human brain contains regions that are not only sensitive to, but also strongly selective to, human voices. This finding is important for several reasons. First, it draws a strong parallel with the architecture of the visual cortex, where face-selective regions have been observed following similar experimental paradigms^{8,9}. It strengthens the view that the different sensory cortices could share common principles of functional organization. Second, it could lead to new comparisons between species, by suggesting that areas sensitive to species-typical vocalizations could be found in the homologous regions in other primates. Indeed, language is probably unique to humans, and its possible evolutionary precursors are hard to define and study in other animals. In contrast, we share the ability to reliably extract affective- and identity-related cues from the species-specific vocalizations with many other species, at least of primates^{21,22}. Finally, these data extend the current knowledge on the organization of the human auditory cortex, by identifying regions of the brain involved in the analysis of human voices, a class of auditory objects of high occurrence and ecological interest. □

Methods

Subjects

Fourteen adult subjects (age 22–47 yr), six males and eight females, participated in this study. Six subjects performed expts 1 and 2, six subjects performed expt 3, and two subjects performed all three experiments. They were all right-handed and had normal audition.

Task and stimuli

In all three experiments, subjects were instructed to simply close their eyes and listen to the sounds that would be presented. Auditory stimuli were delivered binaurally at a mean 88–90-dB sound-pressure level, using foam insert earplugs (Etymotic Research) and an MR-compatible pneumatic sound transmission in expts 1 and 2, and electrostatic,

MR-compatible headphones (KOSS) in expt 3. Auditory stimuli consisted of sounds from a variety of sources arranged in 20-s blocks of similar overall energy (RMS) using Mitsyn (WLH) and CoolEdit Pro (Syntrillium Software Corporation). In each experiment, all blocks were composed of sounds from the same number of speakers/sources (12 in expts 1 and 3, 10 in expt 2). Vocal stimuli were obtained from 47 speakers: 7 babies, 12 adults, 23 children and 5 elderly people. In expt 1, stimuli were divided into 21 blocks of vocal sounds only, and 21 of non-vocal sounds only; vocal stimuli within the same block could be either speech (33%: words, non-words, foreign language) or non-speech (67%: laughs, sighs, various onomatopoeia). Sounds that did not involve vocal-fold vibration were excluded (for example, whistling, whispered speech). Non-vocal stimuli consisted of sounds from nature (14%: for example, wind, streams), animals (29%: cries, gallops), the modern human environment (37%: cars, telephones, aeroplanes) or musical instruments (20%: bells, harp, instrumental orchestra). In expt 2, stimuli consisted of 40 blocks divided into 8 blocks for each of 5 classes: bells; human non-vocal sounds (snaps, footsteps); amplitude-modulated white-noise shaped by the same amplitude envelope as the vocal sounds; vocal sounds, divided in 4 blocks of speech sounds, and 4 of non-speech sounds; and scrambled vocal sounds, obtained by randomly intermixing the magnitude and phase of each Fourier component of the vocal stimuli (Fig. 3c). In expt 3, stimuli consisted of 80 500-ms samples of vocal and non-vocal sounds selected among stimuli from expt 1, and either presented in their original form or frequency-filtered through bandpass filters with peaks at 200 Hz (50 Hz bandwidth) or 1,600 Hz (400 Hz bandwidth). After scanning, subjects performed a two-alternative forced-choice task on pairs of stimuli, one vocal and one non-vocal but with the same filtering, where they had to decide which of the two sounds was the vocal sound, and what was the gender of the speaker who produced that sound.

MRI acquisition

Scanning was performed on a 1.5-T Siemens Vision imager. After obtaining a high-resolution T1 anatomical scan, one series of 128 gradient-echo images (TE = 50 ms, head coil, matrix size: 64 $\times$ 64; voxel size: 4 $\times$ 4 $\times$ 5 mm³; 10 slices acquired in the orientation of the Sylvian fissure) of blood-oxygenation-level-dependent (BOLD) signal—an indirect index of neuronal activity—was acquired for each experiment at a 10-s interacquisition interval (21 min 40 s scanning time). The long interacquisition interval ensures low signal contamination by noise artefacts of image acquisition^{23,24}. In all experiments, the 20-s auditory-stimulation blocks were presented in a randomized order with a 10-s silence inter-block interval using Media Control Function (Digivox); the beginning of each block was synchronized with scanning (see Fig. 1a).

Data analysis

BOLD signal images were smoothed (6-mm gaussian kernel), corrected for motion artefact and transformed into standard stereotaxic space²⁵ using in-house software²⁶. Statistical maps were obtained in each individual by computing for each voxel the *t*-value of the Spearman correlation of the voxel's value time-series with an ideal curve representing the desired contrast²⁷. Group-average statistical images were obtained by computing an omnibus test on individual *t*-maps using a pooled estimate of standard deviation²⁸, and a threshold established at *t* = 5.7 (*P* < 0.001), based on the number of resolution elements in the acquisition volume (2,880 resels).

Received 15 September; accepted 18 November 1999.

- Doehring, D. G. & Bartholomew, B. N. Laterality effect in voice recognition. *Neuropsychologia* **9**, 425–430 (1971).
- Papcun, G., Kreiman, J. & Davis, A. Long-term memory for unfamiliar voices. *J. Acoust. Soc. Am.* **85**, 913–925 (1989).
- Van Dommelen, W. A. Acoustic parameters in human speaker recognition. *Lang. Speech* **33**(3), 259–272 (1990).
- Assal, G., Aubert, C. & Buttet, J. Asymétrie cérébrale et reconnaissance de la voix. *Rev. Neurol.* **137**, 255–268 (1981).
- Van Lancker, D. R. & Canter, G. J. Impairment of voice and face recognition in patients with hemispheric damage. *Brain Cogn.* **1**, 185–195 (1982).
- Van Lancker, D. R., Kreiman, J. & Cummings, J. Voice perception deficits: Neuroanatomical correlates of phonagnosia. *J. Clin. Exp. Neuropsychol.* **11**, 665–674 (1989).
- Imaizumi, S. Vocal identification of speaker and emotion activates different brain regions. *NeuroReport* **8**, 2809–2812 (1997).
- Kanwisher, N., McDermott, J. & Chun, M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302–4311 (1997).
- McCarthy, G., Puce, A., Gore, J. C. & Allison, T. Face-specific processing in the human fusiform gyrus. *J. Cogn. Neurosci.* **9**, 605–610 (1997).
- Klatt, D. H. & Klatt, L. C. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.* **87**, 820–857 (1990).
- Zatorre, R. J., Evans, A. C., Meyer, E. & Gjedde, A. Lateralization of phonetic and pitch discrimination in speech processing. *Science* **256**, 846–849 (1992).
- Dehaene, S. et al. Anatomical variability in the cortical representation of first and second language. *NeuroReport* **8**, 3809–3815 (1997).
- Binder, J. R., Frost, J. A. & Bellgowan, P. S. F. Superior temporal sulcus (STS) responses to speech and nonspeech auditory stimuli. *J. Cogn. Neurosci.* **11** (Suppl. 1), 99 (1999).
- Seltzer, B. & Pandya, D. N. Afferent connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res.* **149**, 1–24 (1978).
- Pandya, D. N. Anatomy of the auditory cortex. *Rev. Neurol.* **151**, 486–494 (1995).
- Jones, E. G., Dell'Anna, M. E., Molinari, M., Rausell, E. & Hashikawa, T. Subdivisions of macaque monkey auditory cortex revealed by calcium-binding protein immunoreactivity. *J. Comp. Neurol.* **362**, 153–170 (1995).
- Rauschecker, J. P. Parallel processing in the auditory cortex of primates. *Audiol. Neuro-Otol.* **3**, 86–103 (1998).

18. Kaas, J. H., Hackett, T. A. & Tramo, M. J. Auditory processing in primate cerebral cortex. *Curr. Opin. Neurobiol.* **9**, 154–170 (1999).
19. Mesulam, M. M. From sensation to cognition. *Brain* **121**, 1013–1052 (1998).
20. Ellis, A. W. in *Handbook of Research on Face Processing* (eds Young, A. W. & Ellis, H. D.) 207–215 (Elsevier, Amsterdam, 1989).
21. Watzlawick, P., Beavin, J. H. & Jackson, D. D. in *A Study of Interactional Patterns, Pathologies and Paradoxes* (Norton, New York, 1967).
22. Rendall, D., Owren, M. J. & Rodman, P. S. The role of vocal trace filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations. *J. Acoust. Soc. Am.* **103**, 602–614 (1998).
23. Belin, P., Zatorre, R. J., Hoge, R., Pike, B. & Evans, A. C. Event-related fMRI of the auditory cortex. *NeuroImage* **10**, 417–429 (1999).
24. Hall, D. et al. "Sparse" temporal sampling in auditory fMRI. *Hum. Brain Map* **7**, 213–223 (1999).
25. Talairach, J. & Tournoux, P. *Co-Planar Stereotaxic Atlas of the Human Brain* (Thieme, New York, 1988).
26. Collins, D. L., Neelin, P., Peters, T. M. & Evans, A. C. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr.* **18**, 192–205 (1994).
27. Turner, R. & Jezzard, P. in *Functional Neuroimaging Technical Foundations* (eds Thatcher, R. W., Hallett, M., Zeffiro, T., John, E. R. & Huerta, M.) 69–78 (Academic, San Diego, 1994).
28. Worsley, K. J., Evans, A. C., Marrett, S. & Neelin, P. A three-dimensional statistical analysis for CBF activation studies in human brain. *J. Cereb. Blood Flow Metab.* **12**, 900–918 (1992).

Acknowledgements

We thank S. Milot, P. Bermudez, M. Bouffard, C. Hurst, A. Cormier, G. Leroux, V. Petre and J. Fiedsend for assistance in data acquisition and analysis, T. Paus, A. Evans, M.-H. Grosbras, I. Lussier, J. Hillenbrand, R. Hoge, G. Legoualher, M.-C. Masure, P. Neelin, K. Worsley and Y. Samson for advice, and N. Kanwisher for seminal discussion. This work was supported by France-Télecom, MRC (Canada), McDonnell-Pew, INSERM-FRSQ and NSERC.

Correspondence and requests for materials should be addressed to P.B. (e-mail: pascal@bic.mni.mcgill.ca).

Functional regeneration of sensory axons into the adult spinal cord

Matt S. Ramer*, John V. Priestley† & Stephen B. McMahon*

* Neuroscience Research Centre, Guy's, King's and St. Thomas' School of Biomedical Sciences, St. Thomas' Campus, Lambeth Palace Road, London SE1 7EH, UK

† Neuroscience Section, Division of Biomedical Sciences, Medical Sciences Building, Queen Mary and Westfield College, Mile End Road, London E1 4NS, UK

The arrest of dorsal root axonal regeneration at the transitional zone between the peripheral and central nervous system has been repeatedly described since the early twentieth century¹. Here we show that, with trophic support to damaged sensory axons, this regenerative barrier is surmountable. In adult rats with injured dorsal roots, treatment with nerve growth factor (NGF), neurotrophin-3 (NT3) and glial-cell-line-derived neurotrophic factor (GDNF), but not brain-derived neurotrophic factor (BDNF), resulted in selective regrowth of damaged axons across the dorsal root entry zone and into the spinal cord. Dorsal horn neurons were found to be synaptically driven by peripheral nerve stimulation in rats treated with NGF, NT3 and GDNF, demonstrating functional reconnection. In behavioural studies, rats treated with NGF and GDNF recovered sensitivity to noxious heat and pressure. The observed effects of neurotrophic factors corresponded to their known actions on distinct subpopulations of sensory neurons. Neurotrophic factor treatment may thus serve as a viable treatment in promoting recovery from root avulsion injuries.

In anatomical experiments, rats received complete crush injuries to two or three cervical dorsal roots, and at the same time osmotic minipumps were implanted to deliver neurotrophic factors (NGF, BDNF, NT3, GDNF or vehicle, $n \geq 5$ per factor) intrathecally (12 μ g per day for 7 days). In vehicle-treated animals, few if any axons were spared by the injury, as determined by transganglionic tracing and electrophysiological experiments (see below). The heavy neurofilament protein NF200, the neuropeptide calcitonin gene-related peptide (CGRP) and the purinoceptor P2X₃ were used to identify

large-diameter myelinated and small-diameter unmyelinated peptidergic and non-peptidergic classes, respectively (these three markers specifically identify nearly all dorsal root ganglion (DRG) neurons^{2,3,4}). With rhizotomies of 2 or 3 roots, some CGRP and P2X₃ immunoreactivity remained in the spinal superficial laminae because of collaterals from adjacent, intact roots. However, here we were interested in whether regenerating afferents invaded the CNS portion of the dorsal root. In all vehicle-treated animals, axon terminals halted at the dorsal root entry zone (DREZ); this was verified with glial fibrillary acidic protein (GFAP) immunohistochemistry (Fig. 1a). With NGF treatment, there was a massive invasion of the central side of the entry zone by CGRP-positive fibres, but not by NF200-positive axons. There was also a smaller but significant effect on axons expressing P2X₃ (Fig. 1b). Conversely, BDNF treatment did not induce significant ingrowth of any fibre type, although trkB is present in some large DRG neurons⁵. NT3 promoted a robust regrowth of NF200-expressing axons and, to a lesser extent, axons expressing P2X₃ across the DREZ. GDNF treatment proved to be the most effective, causing significant increases in NF200-, CGRP- and P2X₃-immunoreactive axon densities central to the entry zone. These distinct neuronal phenotypes are known to be responsive to the different neurotrophic factors: some large-diameter neurons possess the high-affinity BDNF receptor trkB, most express trkC⁵, and many express the BDNF receptor components GFR α 1 and RET³; nearly all small CGRP-containing neurons express the NGF-specific receptor trkA⁴, and P2X₃-expressing cells are known to be selectively sensitive to GDNF, implicating expression of receptor components GFR α 1 and RET³. Here we have measured growth across the entry zone, although previous studies^{6,7} indicate that a very small number of damaged axons may circumvent the entry zone by growing along blood vessels. We have also frequently seen an intimate association between regrowing axons and cord vasculature (arrow in lower right panel of Fig. 1a).

When injected into peripheral nerves, the transganglionic tracer cholera toxin fragment B (CTB) is selectively transported to the central terminals of large-diameter afferents in laminae III and IV (Fig. 1c). One week after rhizotomy in vehicle-treated animals, CTB-labelled end-bulbs had regenerated up to, but not beyond the DREZ ($n=5$, Fig. 1c). With trophic factor treatment ($n=5$ per group), CTB-positive terminals were visible central to the DREZ only after NT3 or GDNF treatment. In these animals, ectopic CTB-filled terminals were seen not only in the white matter of the cuneate fasciculus, but also in laminae I and II. These results are consistent with the observed regrowth of NF200-expressing axons in NT3- and GDNF-treated animals.

We determined whether A-fibre electrical stimulation of brachial nerves (500 μ A, 100 μ s) could evoke postsynaptic potentials in the cervical dorsal horn in rhizotomized, NT3-treated rats. In uninjured rats, cord dorsum potentials (average of 32 sweeps delivered at 0.5 Hz) consisted of a stimulus artefact followed by a primary afferent volley and then by a large postsynaptic potential (Fig. 2a). The entire postsynaptic response disappeared after rhizotomy and did not return after 4 weeks of vehicle treatment (Fig. 2b). In NT3-treated animals four weeks post-rhizotomy, peripheral nerve stimulation evoked a modest postsynaptic potential (Fig. 2c) which was abolished by an acute crush (not shown), verifying that it originated from the regrown, rather than an intact root.

Single- and multi-unit activity of dorsal-horn neurons was recorded and poststimulus time histograms (PSTs) were generated from unitary activity at depths of 100 μ m, 500 μ m and 1 mm. Stimulus-driven units could be found in all intact controls (Fig. 2a, middle trace; $n=5$), but in no vehicle-treated rhizotomized rats (Fig. 2b, middle trace; $n=4$). In NT3-treated animals (Fig. 2c, middle trace), 59% of tracks had stimulus-driven units (in each animal: 11/15, 11/15, 7/15, 6/15 and 9/15 tracks. Analysis of the area under the PSTs revealed a significant increase in spike activity driven

by peripheral nerve stimulation in NT3-treated rats over vehicle-treated rats ($P=0.028$, Tukey's post hoc test), although this was still significantly smaller than in uninjured rats ($P=0.004$). When the regrown root was crushed, stimulus-driven units were no longer observed. Figure 2 illustrates the average degree of connectivity seen in these animals, using pseudo-colour plots relating stimulus-evoked activity, time and depth in the dorsal horn (right-hand panels). By four weeks after the operation, evoked activity could be detected at all depths recorded in NT3-treated rats.

To determine whether regenerated axons can make functional connections in the spinal cord following treatment with NGF, BDNF or GDNF, we performed terminal electrophysiological experiments in which brachial nerves were stimulated at C-fibre strength (4 mA, 1 ms), and single- and multi-unit activity was recorded from dorsal-horn neurons (Fig. 3). In vehicle-treated

animals, some units fired spontaneously, but there was no stimulus-driven activity in the dorsal horn. With NGF treatment, there was a significant restitution of stimulus-driven postsynaptic activity ($P<0.001$), mainly at a long latency (30–100 ms), consistent with regeneration of small-diameter afferents. The magnitude of the area under the averaged PSTs was not significantly different from intact controls; this result could reflect either near-complete re-innervation of the dorsal horn, or more probably, increased responsiveness of dorsal-horn neurons to afferent input⁸. Unexpectedly, BDNF treatment resulted in a small but significant recovery of early post-synaptic activity (presumably A-fibre-driven), but not of longer-latency responses. This occurred despite a lack of anatomical evidence for axonal growth across the entry zone (Fig. 1), although again this may have been due to axonal re-entry via blood vessel epithelium^{6,7}. GDNF treatment was also the most effective treat-

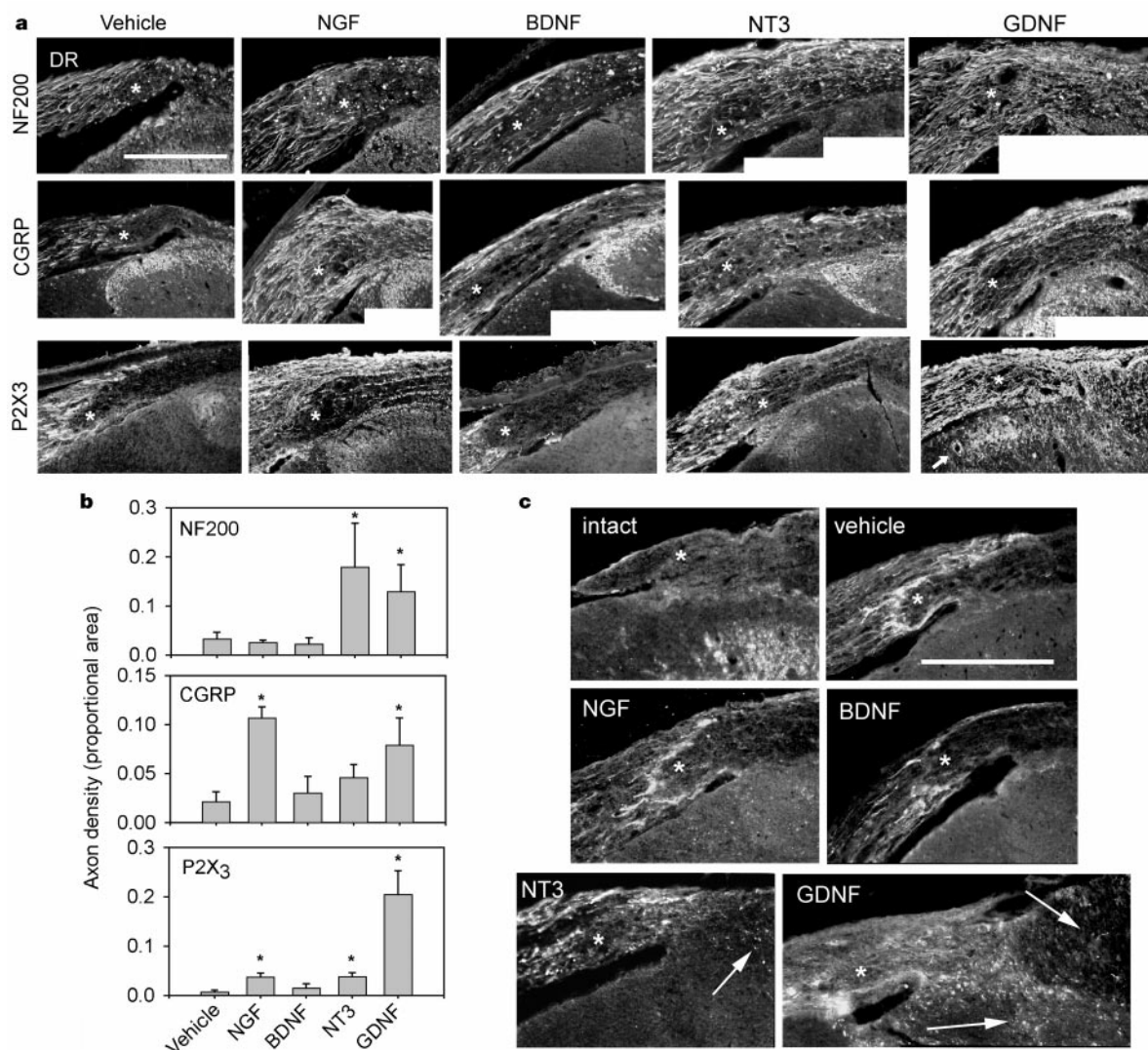


Figure 1 Axonal growth across the dorsal root entry zone, one week post-lesion. In the micrographs shown, dorsal roots (DR in **a**) approach the cord from the left. **a**, In vehicle-treated animals, regenerated axons do not advance beyond the dorsal root entry zone (DREZ). With NGF treatment, many axons expressing CGRP, but few expressing P2X₃ and none expressing NF200, can be seen central to the entry zone. BDNF treatment results in little regrowth of any axonal phenotype across the entry zone, but some CGRP-positive axons can be seen within the CNS portion of the DREZ (not significant (**b**)). NT3 promotes the regrowth of NF200-expressing axons across the DREZ but has little effect on CGRP- or P2X₃-expressing fibres. GDNF promotes the regrowth across the entry zone of all fibre types studied, especially those expressing P2X₃. CGRP immunoreactivity and, to a lesser extent, P2X₃ immunoreactivity is still visible in the lateral part of the dorsal horn because only a small number of dorsal roots (two or three) were crushed, leaving intact longer

roostocaudal collateral branches originating in adjacent roots. In many cases individual ingrowing axons were observed in close association with blood vessels in the dorsal horn (arrow, lower right). **b**, Quantification of axon density in the white-matter portion of the dorsal root, just central to the entry zone. Asterisks indicate significant differences from vehicle treatment. **c**, Transganglionic tracing with CTB. The distribution of CTB-filled axons in intact animals extends ventrally from lamina III, with a complete absence of terminals in lamina I and II. Following rhizotomy, in vehicle-treated animals CTB-filled end bulbs can be seen to have halted at the DREZ. This is also true for NGF- and BDNF-treated animals. With NT3 and GDNF treatment, many CTB-positive terminals can be seen central to the DREZ, in laminae I and II, and in the white matter of the dorsal columns (long arrows). **a**, **c**, All asterisks are placed central to the PNS/CNS border, at the apex of the DREZ (identified by GFAP immunohistochemistry). Scale bars, 200 μ m.

ment in the electrophysiological experiments, resulting in the reappearance of both A-fibre-evoked and C-fibre-evoked responses in the dorsal horn ($P < 0.001$ for both responses). These results are consistent with the regrowth of large (NF200-expressing and CTB-transporting) and small (CGRP- and P2X₃-expressing) axons (Fig. 1).

We also investigated the behavioural consequences of neurotrophic factor treatment in rats with rhizotomies from C4–T2 (to remove all afferent input from the forelimb). Two behavioural tests were used to evaluate recovery of nociception in blind experiments: pinch pressure required to elicit forepaw withdrawal was measured with a Randall-Selitto noxious pressure device (Fig. 4), and withdrawal latency was measured when the forepaw was immersed in a 49°C waterbath (Fig. 4). In all rhizotomized animals ($n = 6$ per treatment group), motor behaviour, although it was imprecise during walking and reaching, appeared otherwise unimpaired, and automatic movements such as grooming appeared completely unimpaired. Septuple rhizotomy plus vehicle treatment resulted in a complete absence of response to either noxious pressure or heat

throughout the three-week postoperative testing period (cut-off latency set at 9.5 s for waterbath, cut-off pressure set at 250 g for Randall-Selitto test). Chronic intrathecal NGF treatment resulted in re-establishment of sensitivity to both thermal and mechanical noxious stimulation after 15 days post-lesion (significantly different

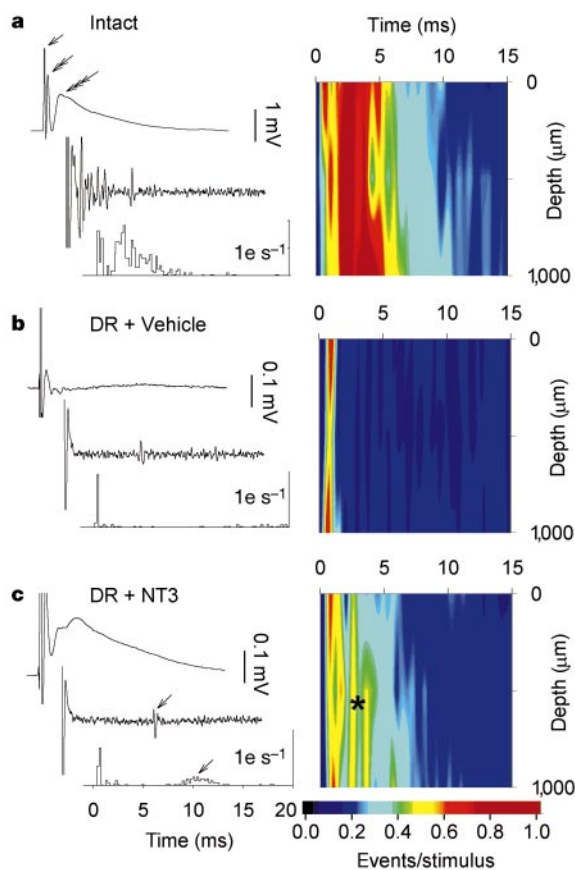


Figure 2 Functional reconnection of regenerating axons (A-fibre stimulation). Shown are cord dorsum potentials recorded with a silver ball electrode (top traces in left-hand panels), single- and multi-unit activity recorded with tungsten microelectrodes (middle traces) and PST histograms (bottom traces), and overall activity represented as a function of time and depth (coloured 'depth profiles' in right-hand panels). **a**, In cord dorsum potentials from intact animals, a stimulus artifact (single arrow) is followed by a primary afferent volley (double arrow) and a postsynaptic potential (triple arrow). Multi-unit recordings and PST-histograms from intact animals show A-fibre-evoked postsynaptic activity time-locked to nerve stimulation. **b**, The large postsynaptic potential is absent in cord dorsum potentials from rhizotomized, vehicle-treated animals, and no stimulus-driven units are apparent in individual PST histograms or in depth profiles. **c**, In NT3-treated rhizotomized animals a small postsynaptic potential reappears in cord dorsum potentials, and stimulus-driven units could be found in many instances. Arrows in **c** indicate a single identified unit and its PST histogram. Differences between all treatments (area under averaged PST histograms) are significantly different ($P < 0.001$, two-way ANOVA, $n = 5$ per group). Colour 'depth profiles' show that in NT3-treated animals, stimulus-driven units could be found at all depths to 1,000 μm.

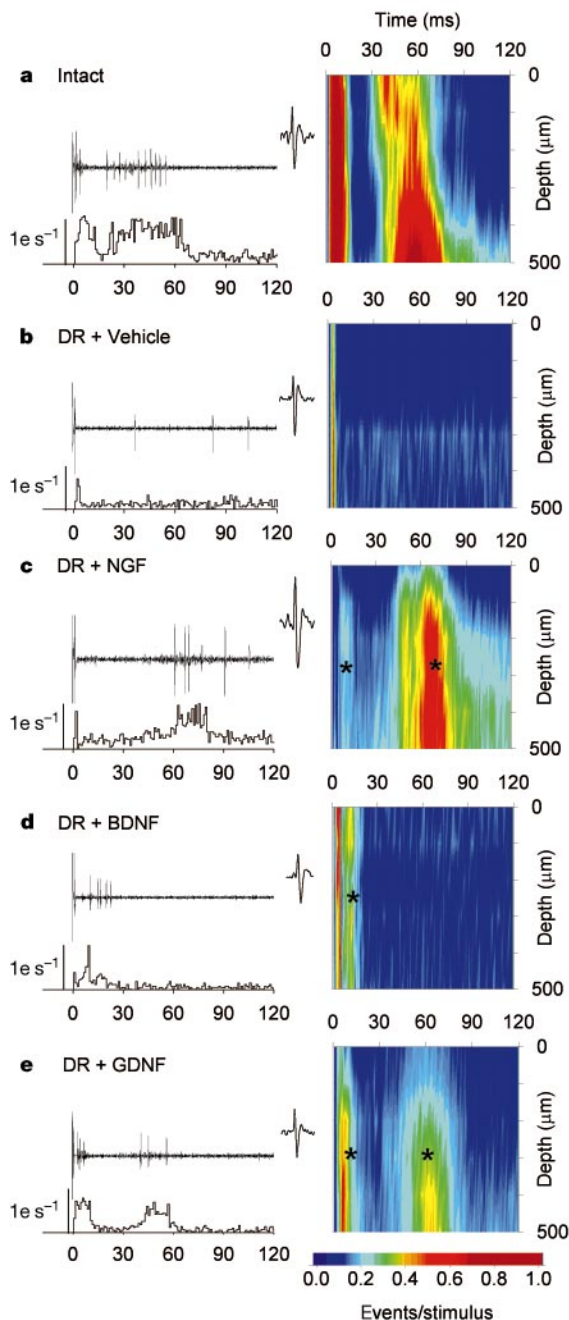


Figure 3 Functional reconnection of regenerating axons (high-intensity stimulation). Sample sweeps of multi-unit recordings and their PST histograms (16 sweeps per histogram) are shown with expanded profiles of single-unit recordings. Also shown are colour 'depth profiles' showing early (A-fibre driven) and late (C-fibre driven) activity as a function of depth. **a**, In intact rats ($n = 5$), early (A-fibre-driven) and late (C-fibre-driven) postsynaptic responses are reliably evoked by supra-threshold, C-fibre-strength peripheral nerve stimulation (4 mA, 1 ms, 0.5 Hz). **b**, In vehicle-treated rats no stimulus-driven activity was observed ($n = 4$ rats). **c**, In NGF-treated animals ($n = 5$) late (presumably C-fibre-driven) activity was regained. **d**, BDNF treatment ($n = 5$) resulted in the reappearance of a small A-fibre-evoked response, but no C-fibre-evoked response. **e**, In GDNF-treated animals ($n = 5$) there was a significant restitution of both A- and C-fibre evoked responses by dorsal horn neurons. Asterisks indicate significant differences at the respective time points from vehicle-treated animals.

from the vehicle-treated group, $P < 0.001$ on both 15 and 20 days postoperative). Neither BDNF nor NT3 treatment had any significant effect on behaviour in either of these tests. GDNF treatment led to a robust recovery of nociception in both tests (significantly decreased withdrawal latency and threshold pressure after 10 days postoperative, maintained throughout the remainder of the experiment).

These results show that chronic intrathecal neurotrophic factor treatment can promote the regeneration of sensory axons across the DREZ and into the spinal cord, cause the reconnection of regenerated axons with neurons in the dorsal horn, and perhaps most importantly, result in the recovery of sensation to noxious peripheral stimuli. These effects have been brought about without the use of peripheral nerve grafts^{9,10}, of transplants of embryonic spinal cord^{11,12} or DRG^{7,13}, or of olfactory ensheathing cells¹⁴, all of which have been employed previously to induce some measure of primary afferent axonal regrowth into the spinal cord. With these techniques, undesirable further damage is incurred as a by-product of reparative surgery. However, others have deliberately used destructive measures such as X-irradiation of the cord to deplete inhibitory central glia¹⁵, and 'conditioning lesions' have been used in attempts to boost regeneration of central processes of sensory neurons by damaging their peripheral branches^{16,17}. Few studies have investigated the effects of supplementary trophic support on dorsal root regeneration into the spinal cord^{10,16,18}, and none has studied functional recovery. The data shown here also comprise the first report of a significant beneficial effect of GDNF on axonal growth into the CNS.

Hypotheses regarding the failure of regeneration into and within the CNS have invoked the prohibitive nature of the CNS environment, often citing the inhibitory properties of myelin and myelin-associated glycoproteins (MAG)¹⁹. Attempts to overcome this inhibition have involved antibodies against neurite growth inhibitors²⁰, or ablation of central glia immunologically²¹ or radiologically²². However, the inhibitory nature of MAG and myelin is not absolute. DRG neurons, when implanted with minimal trauma into white matter tracts, can extend neurites for long distances²³; conditioning lesions also result in limited axonal regrowth in the dorsal columns following spinal cord injury¹⁷; and we have shown that treatment with NT3 can result in growth into and beyond dorsal column lesions²⁴. The conditioning lesion effect may itself be mediated by neurotrophic factors, as messenger RNA for NGF and NT3 (and possibly other factors) is upregulated in satellite cells of the DRG following peripheral nerve injury²⁵.

In vitro, prior exposure of DRG neurons to neurotrophins increases intracellular cyclic AMP and allows axons to grow in the presence of MAG/myelin¹⁹. In the present experiment, DRG neu-

rons would have been similarly 'primed' by neurotrophic factor treatment while axons were regenerating within the roots, before contact with the DREZ. Alternately, neurotrophic factors may have induced a glial response²⁶, facilitating axonal re-entry into the spinal cord. This is less likely, given the remarkable specificity of the action of neurotrophins on appropriate populations of DRG neurons demonstrated here.

The main problem after dorsal root or spinal cord lesions is the lack of axonal elongation within the CNS. In addition, the gliotic response at the DREZ following dorsal rhizotomy mimics that seen following CNS lesions²⁷. However, these models differ in that one involves damage to the peripheral nervous system and one involves direct CNS trauma. Neurotrophins differentially promote regeneration after root or CNS lesions: although injured sensory axons in the dorsal funiculus do regenerate in response to NT3, the reaction of these axons to GDNF is meager²⁴, unlike the present experiment, in which both of these factors induced a robust ingrowth of axons into the spinal cord. Obstructions such as glial scarring and cyst formation, present after CNS injury but not after rhizotomy, may underlie such differences.

It is now evident that with the appropriate support, damaged sensory axons can successfully navigate the DREZ. We believe that GDNF and other neurotrophic factors have vast therapeutic potential in the treatment of dorsal root lesions and, once the nature of central lesions is better understood, of CNS damage in general. □

Methods

Immunohistochemistry and transganglionic tracing

In sodium pentobarbitone-anaesthetized animals (60 mg kg⁻¹), cervical spinal cords were exposed by laminectomy. Two or three roots contributing to forelimb innervation (usually C5–C7) were crushed (three times each, 10–15 s per crush) midway between the DRG and the DREZ. A pre-filled cannula was inserted through the atlanto-occipital membrane along the dorsal surface of the cord until the tip rested at mid-C4. The opposite end was attached to an osmotic minipump (Alzet) containing either NGF (Genentech), NT3 (Genentech), BDNF (Amgen), GDNF (Amgen) or vehicle (1% rat serum albumin (RSA) in sterile saline). Trophic molecules were delivered at a rate of 12 µg per day. In transganglionic tracing experiments, 0.5 µl of 0.5% CTB was injected into the spinal nerve distal to the DRG of the rhizotomized segment. One week post-rhizotomy, animals were killed and perfused with 4% paraformaldehyde in 0.1 M phosphate buffer, and their cords were removed, frozen and sectioned (15 µm) on a cryostat. Sections were processed immunocytochemically for CGRP (host sheep, 1:4,000, Affiniti), P2X₃ (host rabbit, 1:1,000, Roche), NF200 (host mouse, 1:400, N52, Sigma) or CTB (host goat, 1:2,000, List). To delineate the DREZ, in the same sections CNS astrocytes were stained with anti-glial fibrillary acidic protein (GFAP, host rabbit, 1:1,000, Dako). Secondary antibodies were conjugated to either fluorescein isothiocyanate or tetramethylrhodamine isothiocyanate. For P2X₃ staining, the tyramide signal amplification method was used⁴. Quantitative analysis of axon density in a 50-µm-wide strip within the dorsal root along the central side of the entry zone was carried out with the aid of image analysis software: five sections from each animal were randomly selected and axon densities were averaged. Density measurements were compared between groups using a one-way analysis of variance (ANOVA).

Electrophysiology

In terminal electrophysiological experiments, urethane-anaesthetized (1.5 g kg⁻¹) paralysed (gallamine triethiodide, 60 mg kg⁻¹) and respired rats underwent cervical laminectomy to expose the spinal cord from C3 to T3. All dorsal roots were transected unilaterally except one, which had been crushed four weeks previously (usually C6). The major brachial nerves (radial, median, ulnar) were exposed in the upper forelimb and stimulated with silver hook electrodes. In experiments using A-fibre strength stimulation (500 µA, 100 µs, 0.5 Hz), silver ball electrodes were used to record cord dorsum potentials (averaged over 32 sweeps). Additionally, single- and multi-unit activity of dorsal-horn neurons was recorded with a tungsten microelectrode in a series of 15 tracks extending from the cord surface down to a depth of 1 mm (3 rows of 5 tracks parallel to the entry zone, separated by about 100 µm). Post-stimulus time histograms (PSTs) were generated from unitary activity at depths of 100 µm, 500 µm and 1 mm (32 sweeps at 0.5 Hz). In each animal and at each depth the area under PSTs was calculated from 3 to 15 ms post-stimulus (to exclude the stimulus artifact and any primary afferent volley), and areas were compared using a two-way ANOVA.

C-fibre strength stimuli (4 mA, 1 ms, 0.5 Hz) were delivered to brachial nerves with silver hook electrodes. Tungsten microelectrodes were used to record extracellular single- and multi-unit activity from a single row of equally spaced tracks (150–200 µm apart) made along a rostro-caudal line, 50 µm medial to the entry zone. In each track, recordings were made at 100-µm steps from the cord surface. PSTs were constructed from 16 traces of single- and multi-unit activity. These were averaged across tracks at each of six depths (0–500 µm) in each animal, and these averaged data were used for statistical analysis. The area under the A- and C-fibre driven components of the curve was calculated for each animal:

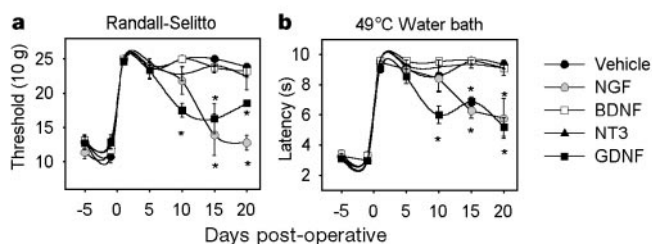


Figure 4 Behavioural recovery following rhizotomy. Functional recovery following dorsal rhizotomy was evaluated with two behavioural tests aimed at assessing responses to noxious stimulation. **a**, Randall-Selitto noxious pressure test. NGF- and GDNF-treated multiply rhizotomized rats showed significant recovery of withdrawal responses by 15 and 10 days post-rhizotomy, respectively. Vehicle-, BDNF- and NT3-treated rats did not respond to noxious pressure throughout the testing period. **b**, Responses to noxious heat using a 49 °C water bath. Again, NGF and GDNF treatment led to recovery of nociception after 15 and 10 days, respectively, whereas vehicle, BDNF and NT3 treatments did not. Asterisks indicate significant differences between neurotrophic factor-treated and vehicle-treated groups on each day post-rhizotomy.

3–29 ms for the early wave and 30–100 ms for the late wave. A two-way ANOVA was used to detect significant differences arising from neurotrophin treatments and depths. In all electrophysiological experiments, after data collection, the remaining dorsal root was crushed acutely to ensure that recorded activity originated from the regenerated segment.

In A- and C-fibre strength stimulation paradigms, colour 'depth profiles' were generated from PST data collected at various depths in the dorsal horn to illustrate stimulus-evoked activity as a function of depth. PST histograms were averaged at each depth across five evenly spaced tracks which ran rostro-caudally, parallel to the entry zone. These were then averaged across all animals in each group, and the averaged response probability was colour-coded.

Behaviour

Dorsal rhizotomies were performed as described above, except that all dorsal roots containing afferent input from the forelimb were crushed (C4–T2). Pain-avoiding paw withdrawal reflexes were tested with a 49 °C water bath and a Randall-Selitto noxious pinch device (UGO). Rats with septuple rhizotomies were lightly restrained and the ipsi- and contralateral forepaws were individually stimulated (by immersion or by pinch). The time until withdrawal or the force of pinch when withdrawal occurred was recorded and averaged with two other trials performed during each testing session. Cutoffs were set at 250 g for noxious pinch and 9.5 s for water-bath immersion. For behavioural experiments, rats with various treatments were randomly assigned to groups of 6 and the testing was done blind. Behavioural responses were compared between treatment groups using a one-way ANOVA.

Received 4 August; accepted 29 October 1999.

- Ramon y Cajal, S. *Degeneration and Regeneration in the Nervous System* (Hafner, New York, 1928).
- Bennett, D. L. H. et al. A distinct subgroup of small DRG cells express GDNF receptor components and GDNF is protective for these neurons following nerve injury. *J. Neurosci.* **18**, 3059–3072 (1998).
- Averill, S., McMahon, S. B., Clary, D. O., Reichardt, L. F. & Priestley, J. V. Immunocytochemical localization of trkA receptors in chemically identified subgroups of adult rat sensory neurons. *Eur. J. Neurosci.* **7**, 1484–1494 (1995).
- Bradbury, E. J., Burnstock, G. & McMahon, S. B. The expression of P2X₃ purinoceptors in sensory neurons: effects of axotomy and glial-derived neurotrophic factor. *Mol. Cell. Neurosci.* **12**, 256–268 (1998).
- McMahon, S. B., Armanini, M. P., Ling, L. H. & Phillips, H. S. Expression and coexpression of Trk receptors in subpopulations of adult primary sensory neurons projecting to identified peripheral targets. *Neuron* **12**, 1161–1171 (1994).
- Chong, M. S., Woolf, C. J., Haque, N. S. & Anderson, P. N. Axonal regeneration from injured dorsal roots into the spinal cord of adult rats. *J. Comp. Neurol.* **410**, 42–54 (1999).
- Kozlova, E. N., Stromberg, I., Bygdeman, M. & Aldskogius, H. Peripherally grafted human foetal dorsal root ganglion cells extend axons into the spinal cord of adult host rats by circumventing dorsal root entry zone astrocytes. *Neuroreport* **6**, 269–272 (1995).
- Thompson, S. W. N., Dray, A., McCarron, K. E., Krause, J. E. & Urban, L. Nerve growth factor induces mechanical allodynia associated with novel A fibre-evoked spinal reflex activity and enhanced neurokinin-1 receptor activation in the rat. *Pain* **62**, 219–231 (1995).
- Houle, J. D., Skinner, R. D., Garcia-Rill, E. & Turner, K. L. Synaptic evoked potentials from regenerating dorsal root axons within fetal spinal cord tissue transplants. *Exp. Neurol.* **139**, 278–290 (1996).
- Oudega, M. & Hagg, T. Neurotrophins promote regeneration of sensory axons in the adult rat spinal cord. *Brain Res.* **818**, 431–438 (1999).
- Itoh, Y. & Tessler, A. Regeneration of adult dorsal root axons into transplants of fetal spinal cord and brain: a comparison of growth and synapse formation in appropriate and inappropriate targets. *J. Comp. Neurol.* **302**, 272–293 (1990).
- Tessler, A., Himes, B. T., Houle, J. D. & Reier, P. J. Regeneration of adult dorsal root axons into transplants of embryonic spinal cord. *J. Comp. Neurol.* **270**, 537–548 (1988).
- Rosario, C. M., Aldskogius, H., Carlstedt, T. & Sidman, R. L. Centrifugal growth in orthotopic grafts of allogenic dorsal root ganglia of adult rats: evidence for possible central ingrowth. *Exp. Neurol.* **115**, 158–162 (1992).
- Ramon-Cueto, A. & Nieto-Sampedro, M. Regeneration into the spinal cord of transected dorsal root axons is promoted by ensheathing glia transplants. *Exp. Neurol.* **127**, 232–244 (1994).
- Sims, T. J. & Gilmore, S. A. Regeneration of dorsal root axons into experimentally altered glial environments in the rat spinal cord. *Exp. Brain Res.* **99**, 25–33 (1994).
- Zhang, Y., Dijkhuizen, P. A., Anderson, P. N., Lieberman, A. R. & Verhaagen, J. NT-3 delivered by an adenoviral vector induces injured dorsal root axons to regenerate into the spinal cord of adult rats. *J. Neurosci. Res.* **15**, 554–562 (1998).
- Neumann, S. & Woolf, C. J. Regeneration of dorsal column fibres into and beyond the lesion site following adult spinal cord injury. *Neuron* **23**, 83–91 (1999).
- Iwaya, K., Mizoi, K., Tessler, A. & Itoh, Y. Neurotrophic agents in fibrin glue mediate adult dorsal root regeneration into spinal cord. *Neurosurgery* **44**, 589–596 (1999).
- Cai, D., Shen, Y., DeBellard, M. E., Tang, S. & Filbin, M. T. Prior exposure to neurotrophins blocks inhibition of axonal regeneration by MAG and myelin via a cAMP-dependent mechanism. *Neuron* **22**, 89–101 (1999).
- Schnell, L., Schneider, R., Kolbeck, R., Barde, Y. A. & Schwab, M. E. Neurotrophin-3 enhances sprouting of corticospinal tract during development and after adult spinal cord lesion. *Nature* **367**, 170–173 (1994).
- Dyer, J. K., Bourke, J. A. & Steeves, J. D. Regeneration of brainstem-spinal axons after lesion and immunological disruption of myelin in adult rat. *Exp. Neurol.* **154**, 12–22 (1998).
- Kalderon, N. & Fuks, Z. Structural recovery in lesioned adult mammalian spinal cord by x-irradiation of the lesion site. *Proc. Natl Acad. Sci. USA* **93**, 11179–11184 (1996).
- Davies, S. J. et al. Regeneration of adult axons in white matter tracts of the central nervous system. *Nature* **390**, 680–683 (1997).
- Bradbury, E. J., Khemani, S., King, V. R., Priestley, J. V. & McMahon, S. B. NT-3 promotes growth of lesioned adult rat sensory neurons ascending in the dorsal columns of the spinal cord. *Eur. J. Neurosci.* **11**, 3873–3883 (1999).

- Zhou, X. F. et al. Satellite cell-derived nerve growth factor and neurotrophin-3 are involved in noradrenergic sprouting in the dorsal root ganglia following peripheral nerve injury in the rat. *Eur. J. Neurosci.* **11**, 1711–1722 (1999).
- Winkler, J. et al. Reversible Schwann cell hyperplasia and sprouting of sensory and sympathetic neurites after intraventricular administration of nerve growth factor. *Ann. Neurol.* **41**, 82–93 (1997).
- Fraher, J. P. The transitional zone and CNS regeneration. *J. Anat.* **194**, 161–182 (1999).

Acknowledgements

We thank E. Bradbury and T. Boucher for critical review of the manuscript, V. Cheah for technical assistance, and D. Shelton and Q. Yan for supplying neurotrophic factors. This work was supported by the European Union (S.B.M. and J.V.P.) and a Prideaux grant from the Special Trustees of St. Thomas' Hospital (S.B.M.). M.S.R. receives a fellowship from the Medical Research Council of Canada.

Correspondence and requests for materials should be addressed to M.S.R. (e-mail: matt.ramer@kcl.ac.uk).

Glutamate release in severe brain ischaemia is mainly by reversed uptake

David J. Rossi^{1,2,3}, Takeo Oshima^{1,2,3} & David Attwell^{1*}

¹ Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK

² Shionogi Company Ltd, 3-1-1 Futaba-cho, Toyonaka 561, Japan

³ These authors contributed equally to this work

The release of glutamate during brain anoxia or ischaemia triggers the death of neurons¹, causing mental or physical handicap. The mechanism of glutamate release is controversial, however. Four release mechanisms have been postulated: vesicular release dependent on external calcium² or Ca²⁺ released from intracellular stores³; release through swelling-activated anion channels⁴; an indomethacin-sensitive process in astrocytes^{5–7}; and reversed operation of glutamate transporters^{8,9}. Here we have mimicked severe ischaemia in hippocampal slices and monitored glutamate release as a receptor-gated current in the CA1 pyramidal cells that are killed preferentially in ischaemic hippocampus. Using blockers of the different release mechanisms, we demonstrate that glutamate release is largely by reversed operation of neuronal glutamate transporters, and that it plays a key role in generating the anoxic depolarization that abolishes information processing in the central nervous system a few minutes after the start of ischaemia. A mathematical model incorporating K⁺ channels, reversible uptake carriers and NMDA (N-methyl-D-aspartate) receptor channels reproduces the main features of the response to ischaemia. Thus, transporter-mediated glutamate homeostasis fails dramatically in ischaemia: instead of removing extracellular glutamate to protect neurons, transporters release glutamate, triggering neuronal death.

We whole-cell clamped area CA1 pyramidal cells to monitor the rise in extracellular glutamate concentration ([glu]_o) produced by ischaemia. To maximize the glutamate-evoked current, cells were clamped at –30 mV to relieve Mg²⁺-block of NMDA receptors. When we switched to an extracellular solution that mimicked ischaemia (see Methods), a series of characteristic changes developed in the membrane current (Fig. 1a). For 5–10 min the current became slightly more inward (67 ± 9 pA; 67 cells). This change was reduced when AP5 (D(-)-amino-5-phosphonopentanoic acid) and NBQX (2,3-dioxo-6-nitro-1,2,3,4-tetrahydrobenzo[f]quinoxaline-7-sulphonamide) were applied to block NMDA and non-NMDA glutamate receptors, respectively (see below) and so is partly due to glutamate release. Then, corresponding to the 'anoxic depolariza-

tion' *in vivo*¹⁰, we observed a rapidly developing large inward current, which reached a peak (the 'anoxic depolarization current', $1,222 \pm 68$ pA, 67 cells) and declined to a less inward plateau. At this time, the Schaffer collateral synaptic current was abolished (not shown). AP5 and NBQX greatly reduced the inward plateau, showing that it was generated largely by the release of glutamate (and/or aspartate) (Fig. 1a). Blocking GABA_A (γ -aminobutyric acid type A) receptors with bicuculline produced only a small additional reduction in current (Fig. 1a); subsequent experiments were performed with bicuculline present throughout, and this did not affect the sequence of current changes (see below). Applying AP5 and NBQX sequentially showed that most ($85 \pm 6\%$ in 9 cells; data not shown) of the maintained glutamate-evoked current was via NMDA receptors (probably because non-NMDA receptors rapidly become desensitized). NMDA receptors mediate most of the Ca^{2+} entry that causes cell death¹, so this current may correlate with subsequent neuronal death.

The anoxic depolarization *in vivo* is associated with a rise in the extracellular potassium concentration¹⁰, $[\text{K}]_o$, to 50–60 mM, but the following data suggest that K^+ influx of itself contributes little to the anoxic depolarization current in Fig. 1a. Applying 30 mM $[\text{K}]_o$ solution to non-ischæmic slices gave only a small inward current (Fig. 2a), whereas 50 mM $[\text{K}]_o$ solution gave a larger inward current which was reduced by AP5 and CNQX (6-cyano-7-nitroquinoxaline-2,3-dione) (Fig. 2b), reflecting depolarization-evoked glutamate release triggered by the larger $[\text{K}]_o$. By contrast, applying glutamate (Fig. 2c) generated a large transient inward current very similar to that evoked by ischaemia in Fig. 1a. These data suggest that most of the ischaemia-induced inward current in Fig. 1a reflects glutamate release rather than K^+ influx.

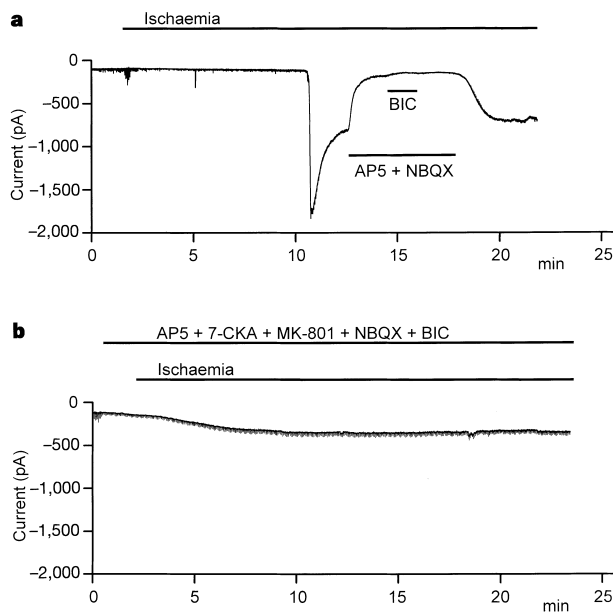


Figure 1 Ischaemia-evoked current changes in pyramidal cells. **a**, After a delay, ischaemia evokes a sudden increase in inward current (anoxic depolarization current) that declines to a plateau. AP5 ($50 \mu\text{M}$) and NBQX ($25 \mu\text{M}$) abolish most of the plateau current. Bicuculline (BIC; $20 \mu\text{M}$) suppresses some of the small remaining current. Adding $50 \mu\text{M}$ MK-801 and $100 \mu\text{M}$ 7-chlorokynurenat (7-CKA) to further block NMDA receptors (3 cells), or 1 mM MCPG ((RS)- α -methyl-4-carboxyphenylglycine) (2 cells) to block metabotropic receptors, gave no further current change. **b**, Response to ischaemia in AP5 ($50 \mu\text{M}$), 7-CKA ($100 \mu\text{M}$), MK-801 ($50 \mu\text{M}$) and NBQX ($25 \mu\text{M}$) to block glutamate receptors, and $20 \mu\text{M}$ BIC. If a sharp anoxic depolarization and $[\text{K}]_o$ rise still occurred in these blockers, a sharp but small inward current shift would occur around $t = 10$ min (perhaps 70–200 pA from the non-glutamate component of Fig. 2a, b). This was not seen in the seven cells tested.

Consistent with this, glutamate-receptor activation was essential for generating the anoxic depolarization current. Cells in slices made ischaemic with NMDA and non-NMDA receptors blocked did not produce an anoxic depolarization current (Fig. 1b); only a slow increase in inward current occurred, like that seen before the anoxic depolarization current in many cells studied without glutamate-receptor blockers (Fig. 3).

By measuring the current blocked by AP5 and NBQX after the anoxic depolarization, we can monitor glutamate release during ischaemia and determine its release mechanism. To examine whether release was Ca^{2+} -dependent, we replaced Ca^{2+} in the superfusion solution with Mg^{2+} (plus 2 mM EGTA to chelate trace Ca^{2+}). This abolished the Schaffer collateral synaptic current, but had little effect on the anoxic depolarization current and subsequent glutamate-mediated inward current (Fig. 3a, g; by contrast, Ca^{2+} removal greatly slows the onset of depolarization in normoxic spreading depression¹¹). Thus, most of the glutamate release in ischaemia is not by $[\text{Ca}^{2+}]_o$ -dependent exocytosis. However, exocytosis may result from Ca^{2+} release from intracellular stores³. We therefore tested the response to ischaemia in zero- $[\text{Ca}^{2+}]_o$ solution containing: dantrolene ($20 \mu\text{M}$), which blocks Ca^{2+} release from stores and ischaemia-evoked miniature excitatory postsynaptic currents³ (Fig. 3b); or thapsigargin ($1 \mu\text{M}$) which blocks Ca^{2+} uptake into stores (Fig. 3c); or thapsigargin ($1.5 \mu\text{M}$) plus caffeine to deplete internal stores of Ca^{2+} (Fig. 3d). None of these procedures significantly reduced the anoxic depolarization current or the subsequent glutamate-mediated current (Fig. 3g). Thus, Ca^{2+} -dependent exocytosis does not contribute much of the glutamate release in ischaemia, probably because ATP and GTP levels fall too low to maintain vesicular release⁹. Previous work suggesting a role for Ca^{2+} -dependent release may have mimicked less severe ischaemia.

A rise of $[\text{Ca}^{2+}]_i$ in astrocytes can release glutamate by a mechanism that is blocked by furosemide⁵ and by the cyclooxygenase blocker indomethacin⁷. The role of this mechanism in ischaemia has not been studied. Furosemide blocks NMDA receptors¹², preventing monitoring of glutamate release. We therefore blocked this mechanism with indomethacin, which affected neither the anoxic depolarization current, nor the subsequent glutamate-mediated current (Fig. 3e, g). Thus, little glutamate is released in ischaemia by this mechanism.

The rise in $[\text{K}]_o$ in ischaemia¹⁰ causes astrocytes to swell¹³, which can lead to glutamate release through swelling-activated anion channels⁴. However, L-644,711, which blocks this process⁴, had no

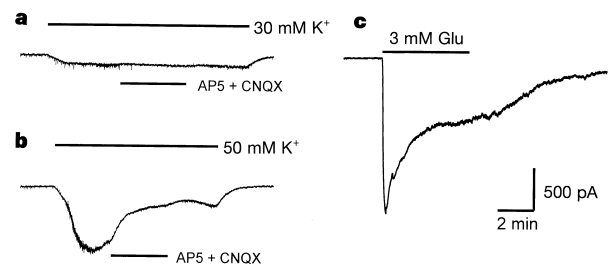


Figure 2 Pyramidal cell response to raised $[\text{K}]_o$ and $[\text{glu}]_o$. **a**, 30 mM K^+ , in non-ischæmic solution with $20 \mu\text{M}$ bicuculline, evokes a small inward current (69 ± 8 pA in 5 cells with Cs^+ as the main pipette cation, as here; 64 ± 12 pA, in 6 cells using a K^+ -based internal; internal Cs^+ blocks K^+ efflux, not influx), with little current due to glutamate release (no suppression by $50 \mu\text{M}$ AP5 and $50 \mu\text{M}$ CNQX). **b**, 50 mM K^+ evokes a larger inward current (832 ± 201 pA, 4 cells), partly through glutamate release. **c**, Glutamate (3 mM) evokes a transient current like ischaemia (peak current $2,199 \pm 355$ pA in 7 cells; a high $[\text{glu}]_o$ is used because uptake reduces its concentration in the slice). NMDA (1 mM , in $25 \mu\text{M}$ NBQX) gave a similar response. The current decay during NMDA superfusion was fitted by an exponential with $\tau = 59 \pm 9$ s ($n = 4$) decaying to $16 \pm 4\%$ of its extrapolated initial value (values for 3 mM glutamate were 46 ± 9 s, $24 \pm 3\%$, $n = 6$).

significant effect on the anoxic depolarization current and subsequent glutamate-mediated current (Fig. 3f, g).

It has been postulated that the rundown of ionic gradients that occurs in ischaemia leads to glutamate release by reversed operation of Na^+ -dependent glutamate transporters (because reversed uptake occurs in cells exposed to ion gradients mimicking ischaemia)^{8,9}. There is no specific non-transported blocker of all glutamate transporters. We therefore tested the role of reversed uptake in ischaemic glutamate release by preloading slices for 1 hour with the glutamate analogue PDC (L-trans-pyrrolidine-2,4-dicarboxylic acid), which is transported four-fold more slowly than glutamate by glutamate transporters¹⁴. When slices are made ischaemic, after external PDC has been washed out, the accumulated intracellular PDC should preferentially occupy the transporters and reduce

glutamate transport out of the cells. During PDC preloading, the external solution contained 1 mM kynurenate to block NMDA and non-NMDA receptors and prevent neuronal damage caused by $[\text{glu}]_o$ rising while transporters were blocked by external PDC (control slices were soaked in kynurenate without PDC). PDC preloading had no effect on the synaptic currents evoked by Schaffer collateral stimulation (Fig. 4a), nor on the cells' response to superfused NMDA or AMPA (α -amino-3-hydroxy-5-methylisoxazole-4-propionate) (Fig. 4b), indicating normal synaptic glutamate release and receptor function. However, PDC preloading produced a three-fold increase in the response to superfused glutamate (Fig. 4b), as would be expected if glutamate transporter activity were reduced,

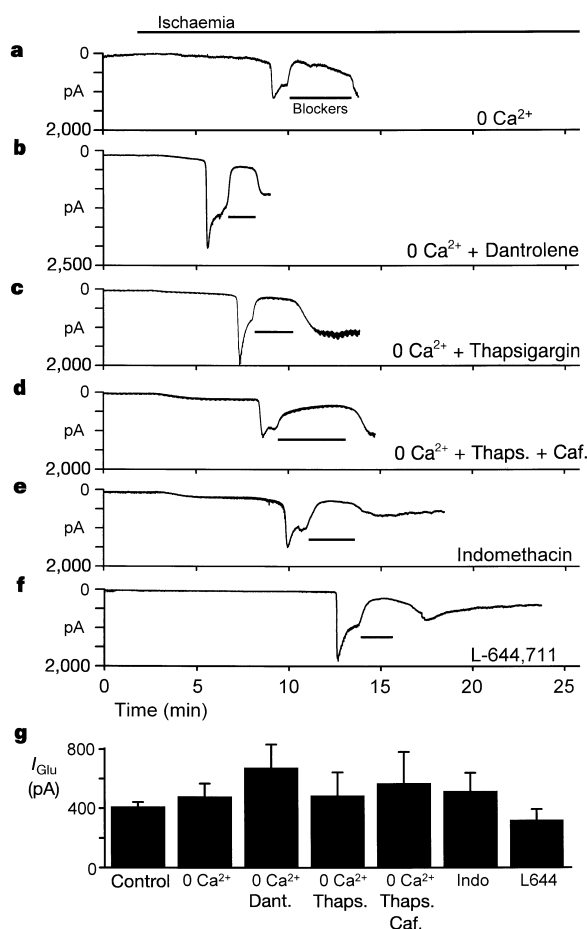


Figure 3 Pharmacology of glutamate release. Drugs were present from 5 min before ischaemia (except thapsigargin (Thaps.) 10 min, and caffeine (Caf.), applied for 2 min at 8 min before ischaemia). Solutions contained BIC (20 μM). **a**, Zero $[\text{Ca}^{2+}]_o$ (+2 mM EGTA). Latency to anoxic depolarization (AD) was less (330 ± 13 s, 25 slices; $P = 6 \times 10^{-7}$) than in normal Ca^{2+} (420 ± 9 s, 67 slices). Tetrodotoxin (1 μM) did not prevent an AD current with normal amplitude and rise time (3 cells, not shown). **b**, Zero $[\text{Ca}^{2+}]_o$ (+2 mM EGTA) plus 20 μM dantrolene (Dant.). **c**, Zero $[\text{Ca}^{2+}]_o$ (+2 mM EGTA) plus 1 μM thapsigargin. **d**, Zero $[\text{Ca}^{2+}]_o$ (+2 mM EGTA) plus 1.5 μM thapsigargin and 10 mM caffeine. **e**, Indomethacin (10 μM). **f**, L-644,711 (300 μM ; blocks swelling-evoked release from astrocytes by 70% with little effect on transporters¹⁴). **g**, Glutamate-mediated current (suppressed by 'blockers' 50 μM AP5 and 25 μM NBQX) 2 min after the AD current, in control solution ($n = 41$ slices) and for the conditions of a–f. P values compared with control were: a, 0.48, $n = 10$ slices; b, 0.69, $n = 11$; c, 0.66, $n = 5$; d, 0.51, $n = 4$; e, 0.44, $n = 11$; f, 0.28, $n = 10$. These manipulations had no effect ($P > 0.05$) on the (non-ischaemic) pyramidal-cell response at -30 mV to 10 μM NMDA and 5 μM AMPA, except that zero Ca^{2+} increased the NMDA response by $26 \pm 5\%$ ($n = 3$), so the glutamate-mediated current in zero Ca^{2+} would be 10% smaller than in control (not significant, $P > 0.55$) if this were taken into account.

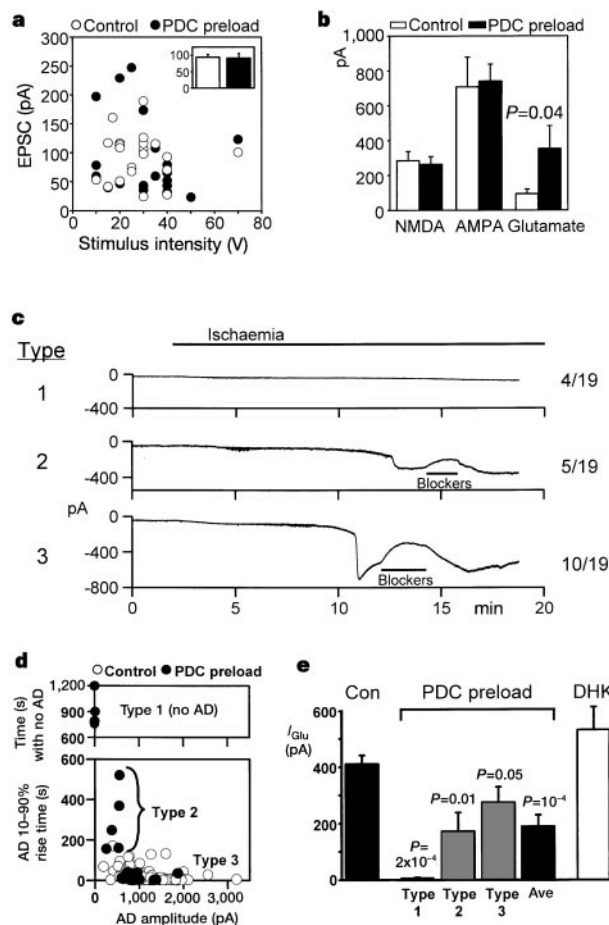


Figure 4 Effects of preloading with the slowly transported glutamate analogue PDC. **a**, Schaffer collateral peak synaptic current (in slices used for ischaemia) as a function of stimulus intensity (each point is 1 cell). Inset, peak current averaged over cells (mean stimulus was 28.1 ± 2.8 V for control, 31.0 ± 3.2 V for PDC; not significantly different, $P = 0.52$). **b**, PDC preload does not affect the pyramidal cell response to NMDA (10 μM ; 7 control and 5 PDC cells) or AMPA (5 μM ; 4 control, 4 PDC cells), nor the pre-ischaemic current at -30 mV (not shown), nor the mean $[\text{glu}]_i$ in the slice³⁰, but increased the response to glutamate (30 μM) three-fold (5 control, 3 PDC cells). **c**, Different responses to ischaemia after PDC preload. Blockers to assess glutamate release were 50 μM AP5 and 25 μM NBQX or 50 μM CNQX. Type 1, no AD current; type 2, small, late AD current (latency 710 ± 122 s, $n = 5$; versus 420 ± 9 s, $n = 67$, in control slices; $P = 1.2 \times 10^{-8}$); type 3, fairly normal AD current (peak $1,012 \pm 127$ pA, $n = 10$; $P = 0.25$ compared with control; latency 472 ± 38 s, slightly prolonged; $P = 0.06$). **d**, Separation of response types after PDC preload (black points). Type 1, no AD; type 2, AD current 10–90% rise time >130 s and peak <600 pA; type 3, 10–90% rise time <100 s and peak >600 pA (overlapping white control points). **e**, Glutamate-mediated current 2 min after the AD current (if present) for control cells (large black bar), for each response type in **c** (shaded bars, P values compared with control), for the average of all responses after PDC-preload (small black bar), and in 300 μM dihydrokainate (DHK, white bar) with no PDC preload.

resulting in less uptake and a higher $[glu]_o$ penetrating into the slice. Preloading with PDC had a striking effect on the pattern of current changes during ischaemia (Fig. 4c, d). In 4 out of 19 slices, the anoxic depolarization current was abolished, and almost no glutamate-mediated current was detectable (Fig. 4c(1), e; this type of response was not seen in any of the 67 slices in normal ischaemic solution: significantly different, $P = 0.002$). In 5 out of 19 slices, an anoxic depolarization current did occur, but it was smaller and slower in onset than normal (Fig. 4c(2), d); the subsequent glutamate-mediated current was less than half of that seen normally (significantly reduced, $P = 0.01$; Fig. 4e). In 10 out of 19 slices, although the anoxic depolarization occurred with approximately the normal time and amplitude of current, the subsequent glutamate-mediated current was about two-thirds of that seen normally (significantly reduced, $P = 0.05$; Fig. 4c(3), e). We attribute the difference in response types to a variable block of reversed uptake by preloaded PDC, possibly through variable loss of PDC from cells between loading and whole-cell clamping, with the type 1 response (no anoxic depolarization or glutamate release) reflecting a more successful block. Averaging over the different response types, the glutamate-mediated current evoked by ischaemia was less than half of that seen normally (Fig. 4e), that is, significantly reduced from the control value ($P = 0.0001$). PDC preloading also reduced glutamate release before the anoxic depolarization ($P = 0.033$; Fig. 5b, c).

Our data suggest that reversed uptake is a major means by which glutamate is released in ischaemia, and plays a crucial role in triggering the anoxic depolarization, consistent with the anoxic depolarization being prevented by blockade of glutamate receptors (Fig. 1b). These observations could be explained if the anoxic depolarization is produced by a positive feedback loop in which metabolic inhibition raises $[K^+]_o$ and $[Na^+]_i$ and depolarizes cells, releasing glutamate, which acts on its receptors to generate further depolarization and K^+ release. Blocking this loop, by blocking glutamate receptors or reversed uptake, would prevent the regenerative rise of $[K^+]_o$ and $[glu]_o$ that causes the anoxic depolarization.

To test this account of events, and to predict the factors shaping the anoxic depolarization, we simulated the cells' response to ischaemia. The cells had three main membrane components (see

Methods): K^+ channels activated by depolarization and $[Ca^{2+}]_i$ (with a small Na^+ conductance to set the resting potential to -80 mV; Fig. 6a), reversible glutamate transporters driven by the prevailing ion gradients (Fig. 6b), and NMDA receptors with a current-voltage ($I-V$) relation shaped by voltage-dependent Mg^{2+} block (Fig. 6c; the inclusion of non-NMDA receptors with rapid and near-complete desensitization¹⁵ had little effect). The resting potential is above the Nernst potential for K^+ , so K^+ ions leave and Na^+ ions enter the cell constantly, but, in normal conditions, they are pumped back by the Na^+/K^+ pump. When ischaemia removes the ATP supply to the pump, K^+ accumulates in the extracellular space (Fig. 6d), depolarizing the cells (Fig. 6e), and the Na^+ gradient driving glutamate uptake is reduced, so that glutamate is moved into the extracellular space (Fig. 6f). When the membrane potential rises above about -60 mV, NMDA receptors activated by the glutamate start to generate current, depolarizing the cells and releasing more K^+ and glutamate (sharp upward inflection at $t = 420$ s on traces labelled 'normal'). Figure 6g shows the current predicted to flow through the sensing cell. Simulations in which glutamate receptors or reversed uptake were blocked gave the traces labelled 'AP5': the sharp rise of $[K^+]_o$ and $[glu]_o$, and the associated anoxic depolarization current, are abolished, as seen experimentally in Fig. 1b and 4c(1). This confirms that a positive feedback loop, in which glutamate release depolarizes cells and releases K^+ by acting on NMDA receptors, can explain the data reported above.

This model groups neurons and glial cells together. Immunocytochemical work^{16,17} suggests that glutamate is released from neurons rather than glia in ischaemia, possibly because their higher¹⁸ internal $[glu]$ will drive reversed uptake more readily. Dihydrokainate ($300 \mu M$), a non-transported blocker of reversed uptake¹⁹ by GLT-1, the main glial transporter in hippocampus, had no effect on the anoxic depolarization current or subsequent glutamate-evoked current (Fig. 4e). Consistent with this, the time course and amplitude of the anoxic depolarization current in hippocampal slices from mice in which the GLT-1 transporter had been knocked out was not significantly different from that in wild-type siblings (manuscript in preparation). We conclude that glutamate release is by reversed operation of neuronal glutamate transporters. These are probably in glutamatergic presynaptic terminals,

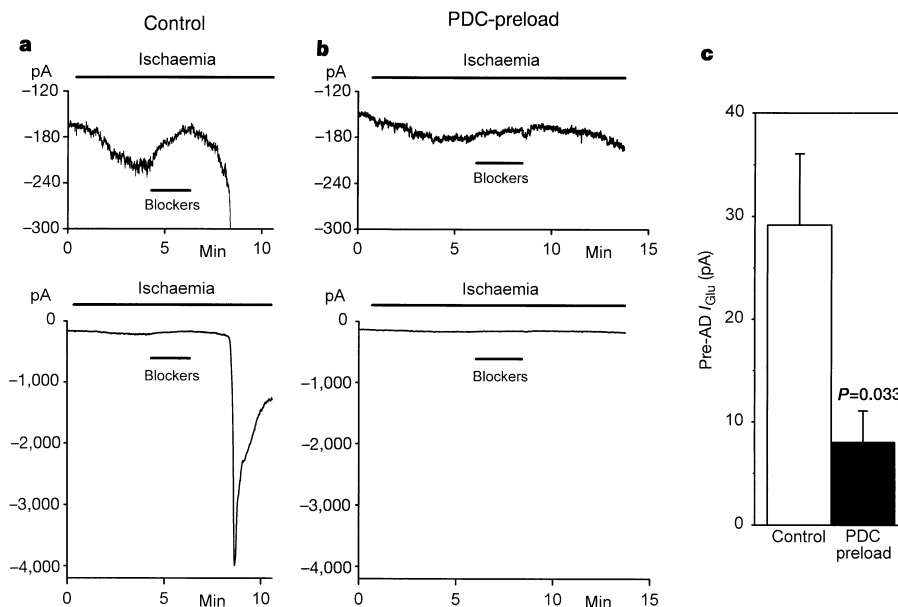


Figure 5 Blocking reversed uptake reduces glutamate release before the anoxic depolarization. Glutamate release was assessed as the current suppressed by $50 \mu M$ AP5 and $25 \mu M$ NBQX (blockers). **a**, Control slice. Top: high gain to show the glutamate-evoked current before the anoxic depolarization. Bottom: low gain to show the anoxic depolarization current. Nine slices gave similar results. **b**, PDC-preloaded slice. Slices

showed a delayed and reduced anoxic depolarization current as in Fig. 4, and 4 of the 10 slices studied showed no anoxic depolarization by 20 min of ischaemia. These slices were not merged with the data of Fig. 4 because they were exposed to blockers before the AD. **c**, Mean current shift evoked by blockers after 6 min ischaemia in control and PDC-preloaded slices.

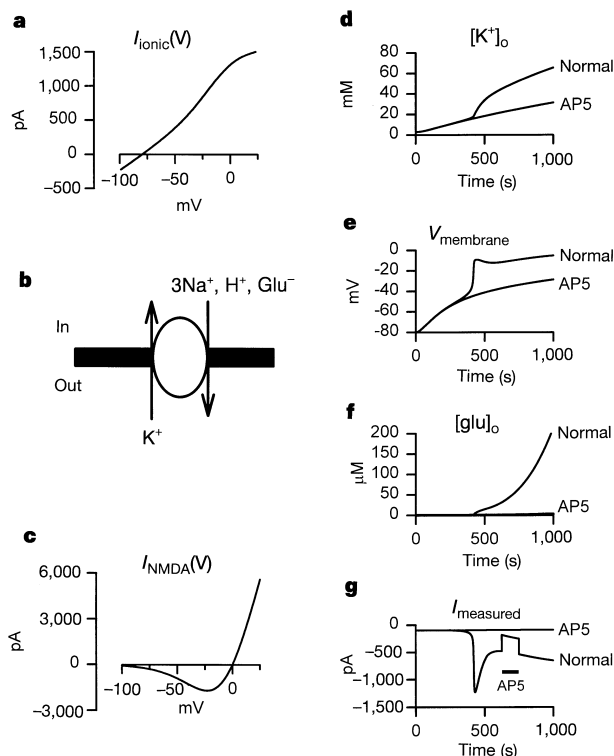


Figure 6 Simulation of the response to ischaemia. **a–c**, Main components of the model; **d–g**, predictions of the model. **a**, Pre-ischaemic I – V relation of cell (excluding Ca^{2+} -gated K^+ current and NMDA channels). **b**, Stoichiometry of reversed glutamate uptake^{19,21}. **c**, I – V relation of fully activated NMDA channels, with Mg^{2+} block at negative potentials. The (normal) plots of $[\text{K}^+]_o$ (**d**), V_{membrane} (**e**), $[\text{glu}]_o$ (**f**) and membrane current of the sensing cell (**g**) were obtained as in Methods. To mimic block of glutamate receptors or reversed uptake, the NMDA conductance was set to zero (curves labelled AP5). Black bar (AP5) on the normal trace in **g** simulates application of receptor blockers as in Fig. 1a, by deleting the cells' NMDA conductance. The NMDA receptor mediated current after 6 min ischaemia (before the anoxic depolarization) is predicted to be 25.4 pA, similar to the 29.2 ± 6.9 pA found experimentally in Fig. 5c.

which have the highest density of neuronal transporters²⁰, and where there may be a locally high internal glutamate concentration helping to drive reversed uptake if synaptic vesicles lose their glutamate to the cytoplasm when ATP levels fall in ischaemia.

It has been suggested that glutamate transporters may continue to take up glutamate during ischaemia²¹, which would reduce neuronal death. By contrast, the present experiments demonstrate that transporter-mediated $[\text{glu}]_o$ homeostasis fails dramatically in ischaemia, with transporters being the major source of the extracellular glutamate that triggers the death of neurons. □

Methods

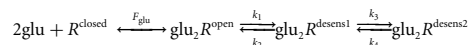
Electrophysiology

Experiments were at 31 °C, on 225- μm -thick hippocampal slices from 12-day-old rats. Pipette solution contained (mM): CsCl 135, NaCl 4, CaCl_2 0.5, HEPES 10, Na_2EGTA 5, MgATP 2, Na_2GTP 0.5, QX-314 10, pH set to 7.1 with CsOH. External solution contained (mM): NaCl 126, NaHCO_3 24, NaH_2PO_4 1, KCl 2.5, MgCl_2 2, glucose 10, bubbled with 95% O_2 /5% CO_2 , pH 7.4. Ischaemia was simulated by replacing 10 mM glucose with 7 mM sucrose, bubbling with 95% N_2 /5% CO_2 and, to get a reproducible fast onset to the ischaemic response, adding 2 mM iodoacetate and 1 mM cyanide (to block glycolysis and oxidative phosphorylation). In experiments raising $[\text{K}^+]_o$, KCl replaced NaCl. Pipette series resistance was ~ 5 M Ω . Data are mean $\pm$ s.e.m. P values are from 2-tailed t - or Fisher's exact test, as appropriate.

Simulation of the response to ischaemia

We solved the differential equations for the membrane potential (V), extracellular potassium ($[\text{K}^+]_o$) and glutamate ($[\text{glu}]_o$) concentrations, and NMDA-receptor desensi-

tization. The hippocampus was treated as a set of identical cells, expressing transporters with the stoichiometry^{19,21} of GLT-1 and EAAT3/EAAC1 (Fig. 6b). The equilibrium extracellular glutamate concentration, $[\text{glu}]_{o,\text{eq}}$, was calculated from^{19,21}: $[\text{glu}]_{o,\text{eq}} = [\text{glu}]_i ([\text{Na}^+]_i/[\text{Na}^+]_o)^2 ([\text{H}^+]_i/[\text{H}^+]_o) ([\text{K}^+]_o/[\text{K}^+]_i) \exp(2VF/RT)$, with $[\text{glu}]_i$ set¹⁸ to 3 mM, $[\text{H}^+]_i/[\text{H}^+]_o$ fixed at 2, and where F , R and T are the Faraday, the gas constant and the temperature, respectively. The transporters were assumed to make $[\text{glu}]_o$ approach $[\text{glu}]_{o,\text{eq}}$ with first-order kinetics ($d[\text{glu}]_o/dt = ([\text{glu}]_{o,\text{eq}} - [\text{glu}]_o)/\tau_{\text{glu}}$) with τ_{glu} (time constant) = 3 s (from¹⁸ salamander retina, corrected for 2 charges moving per glutamate, and to 31 °C with a Q_{10} (factor of increase per 10 °C) of 3; setting $\tau_{\text{glu}} = 42$ s to mimic the effect of acid pH on reversed uptake in ischaemia²² slowed the 10–90% rise of the current in Fig. 6g from 16 to 53 s (experimental value = 27 ± 5 s)). The cell membrane contained an ohmic leak K^+ conductance $G_{\text{K},\text{V}} (= 10 \text{ nS})$, a voltage-activated K^+ current with steady-state conductance $G_{\text{K},\text{V}}(1 + e^{-(4.8+V)/13.6})^{-1} (1 + e^{-(V+67.7)/26.6})^{-1}$ as in hippocampal neurons²³ (with V in mV and $G_{\text{K},\text{V}} = 10 \text{ G}_\Omega$; for the slow changes simulated, gating of this current was at equilibrium), a Ca^{2+} -activated K^+ current with maximum conductance $G_{\text{K},\text{Ca}} = 0.4 G_{\text{K},\text{V}}$ (giving a 28% decrease in input resistance²⁴ at -80 mV) activated in ischaemia with a time course $(1 - e^{-t/\tau})^5$ (fitted to Fig. 4 of ref. 24, with $\tau = 44$ s when adjusted to 31 °C with a Q_{10} of 3), and a leak Na^+ conductance of 0.13 $G_{\text{K},\text{V}}$ (giving a resting potential of -80 mV; Fig. 6a). The cells also had NMDA channels, with a maximum conductance of 25 $G_{\text{K},\text{V}}$, reduced²⁵ by Mg^{2+} block (Fig. 6c) by a factor $(1 + 0.56e^{-0.062V})^{-1}$. The ratio of NMDA conductance to total K^+ conductance was set to produce an initial anoxic depolarization to -10 mV. The NMDA conductance was two-thirds due to Na^+ and one-third due to K^+ , to reproduce a normal reversal potential of 0 mV. Desensitization was modelled by the scheme (equivalent, for slow changes, to desensitization from a closed state with 2 glutamate bound (not shown) which is in rapid equilibrium with the open state)



where the first reaction is at equilibrium with²⁶ $R^{\text{open}}/(R^{\text{open}} + R^{\text{closed}}) = F_{\text{glu}} = 1/(1 + (2.3 \mu\text{M}/[\text{glu}]^{1.5}))$, and $k_1 = 3.94 \text{ s}^{-1}$, $k_2 = 1.94 \text{ s}^{-1}$, $k_3 = 0.0213 \text{ s}^{-1}$ and $k_4 = 0.00277 \text{ s}^{-1}$. These values of k_1 and k_2 result, for a step change of $[\text{glu}]_o$ from zero to a saturating level, in initial desensitization to 33% of the initial conductance¹⁵ with a time constant of 0.17 s (the 0.46 s of ref. 15 adjusted to 31 °C with a Q_{10} of 3). The values of k_3 and k_4 result in subsequent desensitization to 16% of the conductance reached after the fast desensitization, with a time constant of 59 s (see Fig. 2c legend). The membrane potential was calculated from $-C dV/dt = (\text{net membrane ionic current})$, with the membrane capacitance $C = G_{\text{K}} \times 0.1$ s (value not critical, as the changes simulated are much slower than the membrane time constant).

Before ischaemia, K^+ efflux through ion channels is balanced by uptake through Na^+/K^+ pumps. In ischaemia, the activity of the pumps was assumed to decline exponentially ($\tau = 56$ s, that is, the 32 s seen for $[\text{ATP}]$ in hippocampus²⁷, corrected to 31 °C with a Q_{10} of 3), raising $[\text{K}^+]_o$ according to $d[\text{K}^+]_o/dt = (\text{K}^+ \text{ efflux through K}^+ \text{ and NMDA channels minus pumped K}^+ \text{ influx})/(\text{extracellular volume})$. Changes in the Na^+ gradient help to reverse glutamate uptake. $[\text{Na}^+]_o$ was calculated as $143 \text{ mM} - [\text{K}^+]_o$. $[\text{Na}^+]_i$ was initially set²⁸ to 25 mM and $[\text{K}^+]_i$ to 118 mM, and their changes were calculated using an extracellular volume fraction²⁹ of 13% (so they change by $-1/6.7$ the change of $[\text{Na}^+]_o$ and $[\text{K}^+]_o$; values of 5–20% gave similar results). Whole-cell clamping of the sensing cell was assumed not to perturb the $[\text{K}^+]_o$ and $[\text{glu}]_o$ changes; its membrane current was calculated assuming that internal Cs^+ abolished ion efflux (but not influx; see Fig. 2 legend) through G_{K} , $G_{\text{K},\text{V}}$ and $G_{\text{K},\text{Ca}}$. To mimic block of glutamate receptors or glutamate release, the NMDA conductance was set to zero. A regenerative depolarization and increase of $[\text{K}^+]_o$ and $[\text{glu}]_o$, blocked by omitting NMDA receptors, was a robust prediction even with conductances doubled or halved. We set the extracellular volume ($\times$ Faraday) to 3,650 pA/(mM s⁻¹) to make the anoxic depolarization occur 7 min into ischaemia, as seen experimentally. The model can predict the factors shaping the anoxic depolarization. Increasing the ratio of NMDA conductance to $G_{\text{K},\text{V}}$ increased the steady-state $[\text{K}^+]_o$ and $[\text{glu}]_o$, reached and made the anoxic depolarization occur earlier. A faster $[\text{K}^+]_o$ rise could be obtained by increasing the NMDA conductance and making NMDA receptors desensitize more completely. Omitting the Ca^{2+} -activated K^+ current shortened the time to anoxic depolarization (by 40 s), as seen experimentally when external Ca^{2+} was removed (shortened by 90 s), whereas if exocytosis were the main release mechanism Ca^{2+} removal would be predicted to delay the anoxic depolarization.

Received 13 September; accepted 21 October 1999.

- Choi, D. W. & Rothman, S. M. The role of glutamate neurotoxicity in hypoxic-ischemic neuronal death. *Annu. Rev. Neurosci.* **13**, 171–182 (1990).
- Dreier, J., Benveniste, H., Diemer, N. H. & Schousboe, A. Cellular origin of ischemia-induced glutamate release from brain tissue in vivo and in vitro. *J. Neurochem.* **45**, 145–151 (1985).
- Katchman, A. N. & Herschowitz, N. Early anoxia-induced vesicular glutamate release results from mobilization of calcium from intracellular stores. *J. Neurophysiol.* **70**, 1–7 (1993).
- Bednar, M. M., Kohut, J. J., Kimelberg, H. K., Gross, J. J. & Gross, C. E. *In vitro* evidence supporting two mechanisms of action for the anion transport inhibitor L-644,711 in cerebral ischaemia. *Neurosci. Res.* **14**, 53–56 (1992).
- Parpura, V. *et al.* Glutamate-mediated astrocyte–neuron signalling. *Nature* **369**, 744–747 (1994).
- Hassinger, T. D. *et al.* Evidence for glutamate-mediated activation of hippocampal neurons by glial calcium waves. *J. Neurobiol.* **28**, 159–170 (1995).
- Bezzi, P. *et al.* Prostaglandins stimulate calcium-dependent glutamate release in astrocytes. *Nature* **391**, 281–285 (1998).
- Szatkowski, M., Barbour, B. & Attwell, D. Nonvesicular release of glutamate from glial cells by reversed electrogenic glutamate uptake. *Nature* **348**, 443–446 (1990).
- Szatkowski, M. & Attwell, D. Triggering and execution of neuronal death in brain ischaemia: two

- phases of glutamate release by different mechanisms. *Trends Neurosci.* **17**, 359–365 (1994).
10. Hansen, A. J. Effect of anoxia on ion distribution in the brain. *Physiol. Rev.* **65**, 101–148 (1985).
 11. Basarsky, T. A., Feighan, D. & MacVicar, B. A. Glutamate release through volume-activated channels during spreading depression. *J. Neurosci.* **19**, 6439–6445 (1999).
 12. Lerma, J. & Martin del Rio, R. Chloride transport blockers prevent N-methyl-D-aspartate receptor-channel complex activation. *Mol. Pharmacol.* **41**, 217–222 (1992).
 13. Walz, W., Klimaszewski, A. & Paterson, I. A. Glial swelling in ischemia: a hypothesis. *Dev. Neurosci.* **15**, 216–225 (1993).
 14. Sarantis, M. *et al.* Glutamate uptake from the synaptic cleft does not shape the decay of the non-NMDA component of the synaptic current. *Neuron* **11**, 541–549 (1993).
 15. Spruston, N., Jonas, P. & Sakmann, B. Dendritic glutamate receptor channels in rat hippocampal CA3 and CA1 pyramidal neurons. *J. Physiol. (Lond.)* **482**, 325–352 (1995).
 16. Madl, J. E. & Burgesser, K. Adenosine triphosphate depletion reverses sodium-dependent, neuronal uptake of glutamate in rat hippocampal slices. *J. Neurosci.* **13**, 4429–4444 (1993).
 17. Storm-Mathisen, J. *et al.* Ultrastructural immunocytochemical observations on the localization, metabolism and transport of glutamate in normal and ischemic brain tissue. *Prog. Brain Res.* **94**, 225–241 (1992).
 18. Attwell, D., Barbour, B. & Szatkowski, M. Nonvesicular release of neurotransmitter. *Neuron* **11**, 401–407 (1993).
 19. Levy, L. M., Warr, O. & Attwell, D. Stoichiometry of the glial glutamate transporter GLT-1 expressed inducibly in a CHO cell line selected for low endogenous Na⁺-dependent glutamate uptake. *J. Neurosci.* **18**, 9620–9628 (1998).
 20. Gundersen, V., Danbolt, N. C., Ottersen, O. P. & Storm-Mathisen, J. Demonstration of glutamate/aspartate uptake activity in nerve endings by use of antibodies recognizing exogenous D-aspartate. *Neuroscience* **57**, 97–111 (1993).
 21. Zerangue, N. & Kavanaugh, M. P. Flux coupling in a neuronal glutamate transporter. *Nature* **383**, 634–637 (1996).
 22. Billups, B. & Attwell, D. Modulation of non-vesicular glutamate release by pH. *Nature* **379**, 171–174 (1996).
 23. Sah, P., Gibb, A. J. & Gage, P. W. Potassium current activated by depolarization of dissociated neurons from adult guinea pig hippocampus. *J. Gen. Physiol.* **92**, 263–278 (1988).
 24. Nowicky, A. V. & Duchen, M. R. Changes in [Ca²⁺]_i and membrane currents during impaired mitochondrial metabolism in dissociated rat hippocampal neurons. *J. Physiol. (Lond.)* **507**, 131–145 (1998).
 25. Jahr, C. E. & Stevens, C. F. Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics. *J. Neurosci.* **10**, 3178–3182 (1990).
 26. Patneau, D. K. & Mayer, M. L. Structure-activity relationships for amino acid transmitter candidates acting at N-methyl-D-aspartate and quisqualate receptors. *J. Neurosci.* **10**, 2385–2399 (1990).
 27. Raley, K. M. & Lipton, P. NMDA receptor activation accelerates ischemic energy depletion in the hippocampal slice and the demonstration of a threshold for ischemic damage to protein synthesis. *Neurosci. Lett.* **110**, 118–123 (1990).
 28. Friedman, J. E. & Haddad, G. G. Anoxia induces an increase in intracellular sodium in rat central neurons in vitro. *Brain Res.* **663**, 329–334 (1994).
 29. McBain, C. J., Traynelis, S. F. & Dingledine, R. Regional variation of extracellular space in the hippocampus. *Science* **249**, 674–677 (1990).
 30. Roettger, V. & Lipton, P. Mechanism of glutamate release from rat hippocampal slices during in vitro ischemia. *Neuroscience* **75**, 677–685 (1996).

Acknowledgements

We thank E. Rutledge and H. Kimelberg for providing L-644,711, and J. Ashmore, P. Behe, S. Brickley, B. Clark, D. Colquhoun, F. Edwards, M. Farrant, A. Gibb, M. Hamann, M. Häusser and A. Silver for comments on the manuscript. This work was supported by the Wellcome Trust, European Community and the Shionogi Company Ltd.

Correspondence and requests for materials should be addressed to D.A. (e-mail: D.Attwell@ucl.ac.uk).

Blocker protection in the pore of a voltage-gated K⁺ channel and its structural implications

Donato del Camino, Miguel Holmgren, Yi Liu & Gary Yellen

Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA

The structure of the bacterial potassium channel KcsA¹ has provided a framework for understanding the related voltage-gated potassium channels (Kv channels) that are used for signaling in neurons. Opening and closing of these Kv channels (gating) occurs at the intracellular entrance to the pore, and this is also the site at which many open channel blockers affect Kv channels^{2–4}. To learn more about the sites of blocker binding and about the

structure of the open Kv channel, we investigated here the ability of blockers to protect against chemical modification of cysteines introduced at sites in transmembrane segment S6, which contributes to the intracellular entrance. Within the intracellular half of S6 we found an abrupt cessation of protection for both large and small blockers that is inconsistent with the narrow 'inner pore' seen in the KcsA structure. These and other results are most readily explained by supposing that the structure of Kv channels differs from that of the non-voltage-gated bacterial channel by the introduction of a sharp bend in the inner (S6) helices. This bend would occur at a Pro-X-Pro sequence that is highly conserved in Kv channels, near the site of activation gating.

Figure 1 shows the results of a typical blocker protection measurement, for a cysteine introduced at position 474 in the S6

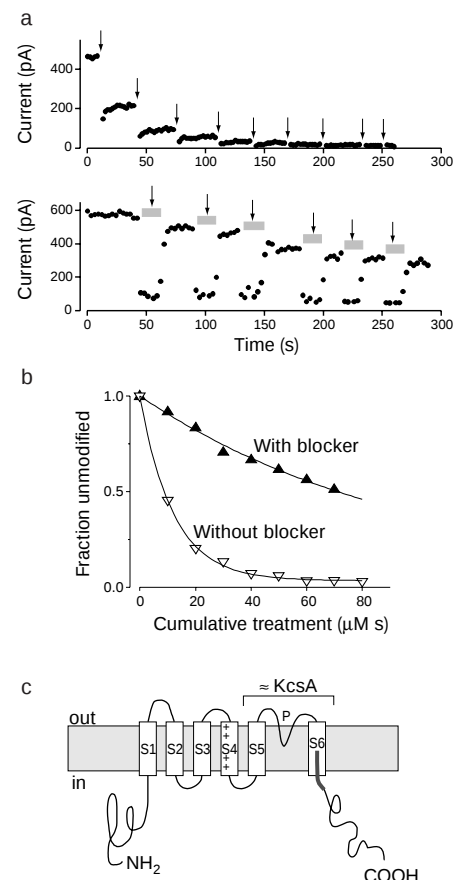


Figure 1 Tetrabutylammonium (TBA) protects Cys 474 from MTSET modification.

a, Chemical modification of 474C mutant channels. Top, the inside-out patch was held at −90 mV; dots indicate the steady-state current in response to brief pulses (20 ms) to +60 mV. At the times indicated by arrows, 20 μM MTSET was applied for 500 ms during a voltage step to +60 mV. Bottom, a similar experiment with modification performed in the presence of 60 μM TBA. The channels were exposed to the blocker during the time showed by the horizontal bars; ~85% of the current was blocked. MTSET (40 μM) was applied for 250 ms in the open state as indicated by the arrows. **b**, Time course of the normalized current decrease as a result of modifications performed without or with 60 μM TBA. The abscissa shows the cumulative reagent exposure (modification time × [MTSET]). The lines represent the best fit of the current reduction data to single exponential functions; the fit for the modification in the presence of blocker was constrained to use the final steady-state reduction seen in the control. The modification rate constant from the exponential fits was ~8.3 × 10⁴ M^{−1}s^{−1} without TBA and ~1.0 × 10⁴ M^{−1}s^{−1} with blocker. Given the degree of blockade, this implies that blocked channels are almost completely unreactive (see Methods). **c**, Model of the membrane topology of a subunit of Shaker K⁺ channel showing the region studied (thick line from positions 470 to 486). As indicated, the S5-P-S6 regions correspond to the sequence homology with the KcsA channel.

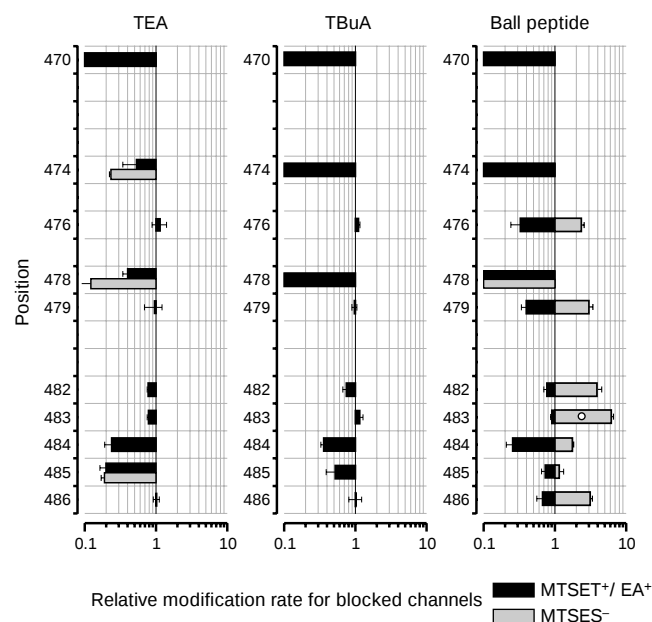


Figure 2 Effects of blocker on MTS reagent modification of cysteines introduced in S6. On a logarithmic axis, the bars represent the change in the modification rates produced by the blockers TEA, TBuA or ball peptide. Bars originate at 1 and end at a position corresponding to the relative modification rate for blocked channels (see Methods). Black bars, positively charged MTSET (or MTSEA, for 470C); grey bars, negatively charged MTSES. Modification at most positions was monitored by the reduction of current (see Methods). The open circle for ball peptide at position 483 shows the value of the relative modification rate for the uncharged reagent MTSACE. Bars that reach to 0.1 represent measured relative modification rates between zero and 0.1. For the modification rates for each residue measured in the absence of blocker, see Supplementary Information.

segment of a Shaker K^+ channel. The control modification rate was determined in an inside-out membrane patch containing many 474C channels (Fig. 1a, top) by repeated brief applications of MTSET (the linear methanethiosulphonate (MTS) derivative of ethyl(trimethylammonium)) to the intracellular surface. After each application there was a step reduction in the size of the current, and these steps were plotted to determine the modification rate constant (Fig. 1b). We performed a similar experiment in the presence of a blocker, tetrabutylammonium (TBuA; Fig. 1a, bottom). The two rate constants were compared (Fig. 1b), and the ratio was corrected for the fraction of channels that remained unblocked (see Methods). In this case, the relative modification rate for blocked channels was near zero.

We systematically applied this approach to determine the effects of three different open channel blockers—tetraethylammonium (TEA), tetrabutylammonium and the inactivation-related 20-amino-acid 'ball peptide'²⁵—on modification of a series of S6 residues. The results are shown in Fig. 2 on a logarithmic scale: the bars to the left represent protection by bound blocker (a decrease in rate) and those to the right represent enhancement of modification by bound blocker.

We first consider results with the positively charged reagents MTSET and MTSEA (the MTS derivative of ethylamine; black bars). The clearest effects were the virtually complete protection ($>10\times$ reduction in modification rate) seen at three positions: 470, 474 and 478. These three positions were fully protected by the two larger blockers, TBuA and ball peptide. The result is consistent with the ideas that these three residues line the pore and that the blockers physically obstruct the pathway from the intracellular solution into the pore. The smaller blocker, TEA, produced complete protection only at the deepest of the three positions (470), indicating that it may penetrate more deeply into the pore. This conclusion is compatible with known differences between TEA and larger quaternary ammonium compounds in sensitivity to mutations⁶ and in allosteric effects on C-type inactivation⁷.

For many positions, the bound blockers had little or no effect on the cysteine modification rate. This was particularly striking in the case of position 476, which lies between two strongly protected residues (474 and 478) but was completely unaffected by the two blockers TEA and TBuA. This residue is thus likely to face away from the pore lining, in a position that allows access from the intracellular solution even when the pore is blocked. This result is consistent with the failure of Cd^{2+} binding or MTSET modification to prevent conduction through channels with a cysteine at this position, even though both treatments completely inhibit the 474C and 478C mutants³.

There are many other effects and non-effects of blockers on modification whose interpretation is less obvious. To the extent that protection occurs by physical occlusion, weak or absent protection could result if the introduced cysteines are completely outside the pore and the binding site for the blockers. Alternatively, weak 'occlusive protection' could occur at a wide part of the pore where the blocker fails to cover simultaneously the cysteines in all four subunits. There are also two alternative mechanisms by which blockers may alter the modification rate: the positive charge on the blocker could repel the positively charged modifying agent (electrostatic), or the blocker could alter the structure of the protein to affect access to the site (allosteric).

The most testable of these is the electrostatic mechanism, which would predict a substantially greater effect for the ball peptide (with charge $q = +6$) than the other blockers (with $q = +1$). If a direct charge-charge interaction is responsible, we should be able to reverse the effect by using a negatively charged modification reagent, MTSES (the MTS derivative of ethylsulphonate). This reversal was seen for the effects of ball peptide at many positions: 476, 479, 482, 483, 484 and 486 (Fig. 2, grey bars). For two of these, 476 and 479, the effect of ball peptide on MTSES modification was equal and opposite to that on MTSET modification, and there was no effect of the two singly charged blockers; this is consistent with a purely electrostatic effect of the ball peptide on modification. At positions 476 and 486 we also found a strong reciprocal effect on ball peptide affinity when the cysteines were modified with charged reagents. Modification with MTSET substantially reduced ball peptide affinity, whereas modification with MTSES increased its affinity (data not shown). These results argue that the electrostatic effect involves a direct interaction between the modified cysteines and the charged ball peptide. Thus the ball peptide must bind physically near these sites, as previously shown for a site in the S4-S5 linker⁸.

The other sites of electrostatic interaction showed asymmetrical effects, which probably represent a combination of electrostatic and other (allosteric or weak occlusive) effects. We tested this idea at position 483 using an uncharged modification reagent, MTSACE (the MTS derivative of propionamide). The effect of ball peptide on modification with this reagent was intermediate between its effects on MTSET and MTSES, and probably represents an allosteric effect on modification at this site. Other, unexplained allosteric effects occurred at position 484 for all three blockers and at position 485 (where the effect was strongest for TEA). (Although the state dependence of cysteine modification³ at these two positions is weak by comparison with the deeper part of the pore, there are 3- to 6-fold changes in accessibility that are compatible with some movement. The known effects of the blockers on gating state do not, however, explain the allosteric protection.)

How do the protection results fit with the known structure of another K^+ channel, KcsA, which is not voltage-gated? The failure of the ball peptide to produce any strong protection below the level of position 478 is unexpected. This result on open Kv channels is apparently inconsistent with the KcsA crystal structure. In KcsA there is a narrow 'inner pore' below the level of the bundle crossing

formed by the inner helices (Fig. 3). If the ball peptide blocked the inner pore at the intracellular entrance, any side chain facing the inner pore would be protected strongly by the blockade: we would thus expect at least one of the residues in the range 482–486 to be strongly protected. The contradiction is most apparent when the pore surface of KcsA is coloured with the protection results from Shaker, and the blockers are viewed on the same scale as the KcsA structure (Fig. 3).

Flexibility or loose packing of the inner pore lining might explain why it remains accessible to reagent even when ball peptide is bound, but it would then become harder to explain why potassium ion flow is blocked by the ball. It is possible that the peptide binding produces a remote (allosteric) action that closes the pore to potassium ions at a higher level (around 478), but there is evidence that the ball acts by physically occluding the pore^{5,9,10}. In particular, TEA and the ball peptide act competitively⁹, an observation most simply consistent with pore blockade. The kinetics of ball peptide action⁵ and the ability of elevated extracellular $[K^+]$ to speed recovery from inactivation¹⁰ also support the idea that the ball peptide acts by directly occluding the pore.

An alternative model to explain the protection results is that in the Kv channels the S6 helix is bent near the intracellular surface of the membrane (Fig. 4), unlike the linear inner helices of KcsA. This would be a natural consequence of the sequence Pro-X-Pro at positions 473–475, which is absolutely conserved among all Kv

channels but is not found in KcsA or in Kir channels. In known protein structures, single prolines generally introduce a kink of about 26° in a helix, because they eliminate a backbone hydrogen bond¹¹. The sequence PxP, which would break two hydrogen bonds within one helical turn, almost never occurs within a helix except within the first turn¹¹ (see Methods). In general, the PxP sequence produces a complete disruption of helical structure, with two helices separated by a break of about four or more nonhelical residues that permit a turn of almost any angle. Of course, initial attempts at modelling the Kv channel structure predicted a bend in S6 at the PxP sequence¹² (the lower S6 was proposed to remain roughly perpendicular to the membrane). But since the publication of the KcsA structure¹ and its good agreement with the functional work on Shaker^{3,4,13}, it has been widely assumed that the KcsA linear inner helices were a good model for the Kv channel S6 structure.

The 'bent S6' model of Fig. 4 fits well with the protection results because below the point of the bend (assumed to span roughly from 473 to 476) the pore would become quite broad. The ball peptide could then bind to this broad face, covering the entrance but not aligned with the centre (analogous to the binding of agitoxin to the extracellular entrance of Shaker channels¹⁴). There would be few positions that exhibit strong protection—because the ball peptide would not cover all four subunits at once—but many positions that could interact with the peptide electrostatically.

In a mutant Shaker channel with a cysteine at position 476, Cd^{2+} ions can lock the channel open by forming an intersubunit bridge between this cysteine and the histidine at position 486 in the neighbouring subunit¹³. Qualitatively this result fit well with the 'bundle crossing' seen in the KcsA structure: the intersubunit bridge would stabilize the bundle in the open configuration, whereas during normal closing the bundle would rearrange and pinch the channel shut at its narrowest point^{3,4}. Quantitatively, though, these positions are located too far apart in the KcsA structure for a bridge to form: based on the position of the KcsA backbone, the S_y-N_ϵ distance would be at least 13 Å (Fig. 4, left), compared with a distance in zinc finger proteins of ~ 3.6 Å (see, for example, ref. 15). The 'bent S6' model of Fig. 4 would also resolve this problem, allowing the bridge to form between neighbouring subunits (the hypothetical positions of the Cys- Cd^{2+} -His bridges are shown in the Figure). Moreover, it allows direct access from the intracellular

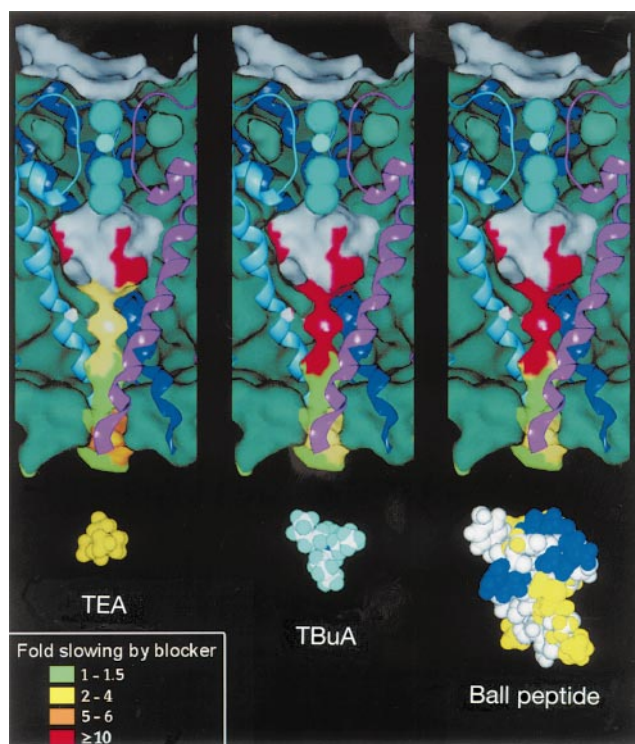


Figure 3 Protection results for three blockers overlaid on the lining of the KcsA pore. Each view shows the KcsA skeleton, with the inner helices crossing near the bottom (one of the four subunits has been removed to permit viewing). The selectivity filter loop is at the top, with three K^+ -binding positions (large cyan balls) and one water molecule (small cyan ball) indicated. The protein surface is a grey shell (dark green from the interior), and the surfaces of the central cavity and inner pore are coloured using the MTSET and MTSEA protection results from Fig. 2. Below the protection pattern for each blocker is a space-filling structure of the blocker. In the case of ball peptide, the first 20 residues of one of the nuclear magnetic resonance structures for the Kv3.4 ball²⁷ are shown, with the basic residues coloured dark blue. The blockers are shown for approximate size and not for exact structure, which is probably mobile. However, even if the bound ball peptide were to adopt a completely different structure, there is too much mass for it to fail to protect the lower S6 if it acts by plugging a pore with the approximate structure of KcsA. The Figure was generated using the program SPOCK (J. Christopher, Texas A & M University).

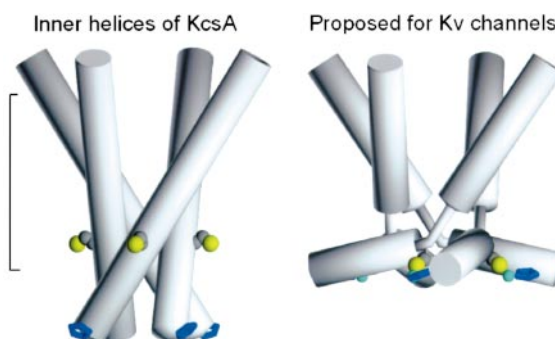


Figure 4 The 'bent S6' model for Kv channels compared with the inner helices of KcsA. The positions of the KcsA inner helices are approximated by four cylinders, one from each subunit. The square bracket next to KcsA shows the approximate extent of the membrane, as judged from the positions of tryptophan layers in the KcsA structure. For Kv channels, the S6 segments are shown as two helices with an intervening bend, proposed to occur near the two prolines at positions 473 and 475. The intracellular entrance is at the bottom, and the bend produces a broad vestibule just below the bundle crossing. The hypothetical position of the intersubunit bridges Cys476- Cd^{2+} -His486 are shown for the Kv channels (yellow ball, cysteine sulphur; cyan sphere, Cd^{2+} ; blue pentagon, histidine imidazole ring). For comparison, cysteine and histidine have been shown at the analogous position in KcsA. This Figure was derived from the KcsA structure¹ and from an exploratory molecular model of the Kv structure (prepared with Insight 98.0 (Molecular Simulations) and exported to a drawing program).

surface to the cysteine at 476; previously it was necessary to suppose that there was a crevice on the back side of the pore-lining helices to permit such access.

This 'bent S6' model for voltage-gated K⁺ channels is as faithful as possible to the one known structure of a related protein, KcsA. The addition of a bend is based on the behaviour in known protein structures of the Pro-X-Pro sequence, which is conserved absolutely among voltage-gated K⁺ channels and is not found in KcsA. The change is mandated by the qualitative constraint on the Kv pore structure imposed by the protection results—the pore becomes broad below the bundle crossing—and by the quantitative distance constraint imposed by the 476C-Cd²⁺-H486 intersubunit bridge. The resulting picture provides an improved basis for further functional work on Kv channels.

The protection data and the Cd²⁺ bridge pertain only to the open state of Kv channels, so it is possible that the Kv channels adopt a KcsA-like structure when they close. Indeed, for KcsA there is evidence that pH-induced gating involves a separation of the lower part of the inner helices¹⁶. Because the spin-labelling data do not permit a quantitative estimate of the changes in distance, it cannot be ruled out that the inner helices may splay out as we propose for the Kv channels, rather than moving as a rigid body as proposed in ref. 16. However, the idea that the Kv channels close to a KcsA-like structure with a narrow inner pore does not seem compatible with the small state-dependent accessibility seen for the lower S6 helix³, even for reagents larger than MTSET (D.D.C. and G.Y., unpublished data). We therefore suspect that the Kv channels have a bent S6 in both the open and closed states. Because the bend occurs quite near the site at which activation gating closes the pore³, it seems possible that gating involves a structural change at the bend. □

Methods

Mutagenesis and expression

We introduced mutations into a Shaker H4 potassium channel¹⁷ containing three additional modifications: a deletion between residues 6 and 26 that removes N-type inactivation¹⁸, substitution of cysteines 301 and 308 by serines to diminish the wild-type effect of chemical modification³, and mutation of threonine 449 to valine to reduce C-type inactivation¹⁹. Channels were transiently expressed in human embryonic kidney (HEK)293 cells and used 1–3 days after transfection. Methods for mutagenesis, transfection and identification of transfected cells have been described²⁰. The tandem dimers were subcloned into a construct prepared by L. Heginbotham and R. MacKinnon²¹.

Physiological recording

All experiments were performed on inside-out excised patches²². Standard experimental solutions contained, in mM: 60 NaCl, 100 KCl, 3 CaCl₂, 1 MgCl₂ and 10 HEPES, at pH 7.4 (pipette); and 160 KCl, 0.5 MgCl₂, 1 EGTA and 10 HEPES, at pH 7.4 (bath). Solutions with Cd²⁺ contained no EGTA. The methods for electrophysiological recordings and rapid perfusion switches have been described^{3,8}.

Synthetic ball peptides

Unless stated otherwise, 'ball peptide' experiments were performed with the E12K, D13K ball peptide²³, which is a double mutant of the 20-amino-acid Shaker inactivation peptide⁵ with increased positive charge (two positively charged residues replace two negatively charged residues). In several cases, we compared the effects of the wild-type Shaker ball peptide. Both ball peptides were synthesized, purified by HPLC and then amidated at the carboxy terminus by the Massachusetts General Hospital peptide synthesis facilities.

Chemical modification and analysis of blocker effects

MTS reagents were purchased from Toronto Research Chemicals Inc. These reagents were prepared as stocks in water each day, held on ice and mixed into the recording solution immediately before use. At most positions, we used the reduction in current to monitor modification, but when this change was small or absent we used other methods. At position 476 we measured the fraction of current with altered gating. The loss of Cd²⁺ sensitivity was used at positions 483 and 486 for MTSET, and also at position 484 for MTSES. Current increase was measured at positions 482, 483 and 486 for MTSES. Positions 471, 472, 473 and 480 were not substantially affected by MTSET or MTSEA treatment. 479C mutant channels were studied as tandem dimers, with only one of the two protomers containing the mutation (this mutant did not express functionally as homotetramers); the F481C mutant did not express even as a tandem dimer. The cysteine mutants at position 475 (normally prolines) were not included in this study because the mutant itself has pronounced effects on gating and blockade. The cysteine at position 477 has a much lower reaction rate than its neighbours and was also not included. MTSEA was the only reagent that produced a measurable effect on 470C. We obtained the reaction

rates by measuring the time course of the effect caused by the modification upon repeated brief application of the MTS reagent under conditions of maximum open probability. We calculated modification rate constants by taking the reciprocal of the time constant of a monoexponential fit to the time course of modification and dividing by [MTS] (Fig. 1; also ref. 3). The effect of the blockers was assessed by measuring the modification rate constant in the presence or absence of the blocker, using control experiments done with the same MTS stock solution on the same day. For the channel blockers used here, it has been established kinetically that a single molecule of blocker produces the blockade, so we can consider that in the presence of the blocker, channels are in two states: unblocked or blocked^{5,24,25}. In a non-saturating concentration of blocker we measured the fraction unblocked (f_{UB}) directly from the ratio of the current in the presence of blocker to the current in its absence. The measured rate constant for modification in the presence of blocker ($k_{overall}$) is then $k_{overall} = f_{UB} \cdot k + (1 - f_{UB}) \cdot k'$, where k and k' are the modification rate constants for unblocked and blocked channels, respectively. The rate constant k was measured directly, and k' was computed from the measurements of k , $k_{overall}$ and f_{UB} . The ratio $r = k'/k$ gives the relative modification rate for the blocked channels (plotted in Fig. 2 for many positions and several reagents). Small values of r correspond to strong protection; $r = 1$ corresponds to no protection; and $r > 1$ means that the blocked channels react faster than the unblocked channels.

Analysis of Pro-X-Pro in the protein data bank

To allow us to check the conclusions of ref. 11 on newer structures, the authors (M. W. MacArthur and J. M. Thornton) performed a search²⁶ for non-redundant structures containing the PxP motif with at least one α -helical residue in the ten residues preceding and one in the nine residues following the first proline, to identify prolines in an α -helical context. Of the ~100 structures selected in this way, almost all have four or more non-helical residues between two helical segments. The 14 structures with the highest assigned helical content around the PxP were examined and found to have breaks between the two helices, with an interhelix angle between ~35° and 180° (a hairpin). One structure, represented by entry 1PHB (cytochrome P450), had a short helix preceding PxP that was bent only ~20° relative to the helix following.

Received 17 August; accepted 6 December 1999.

- Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of potassium conduction and selectivity. *Science* **280**, 69–77 (1998).
- Armstrong, C. M. Interaction of tetraethylammonium ion derivatives with the potassium channels of giant axon. *J. Gen. Physiol.* **58**, 413–437 (1971).
- Liu, Y., Holmgren, M., Jurman, M. E. & Yellen, G. Gated access to the pore of a voltage-dependent K⁺ channel. *Neuron* **19**, 175–184 (1997).
- Yellen, G. The moving parts of voltage-gated ion channels. *Q. Rev. Biophys.* **31**, 239–296 (1998).
- Zagotta, W. N., Hoshi, T. & Aldrich, R. W. Restoration of inactivation in mutants of Shaker K⁺ channels by a peptide derived from ShB. *Science* **250**, 568–571 (1990).
- Choi, K. L., Mossman, C., Aubé, J. & Yellen, G. The internal quaternary ammonium receptor site of Shaker potassium channels. *Neuron* **10**, 533–541 (1993).
- Baukrowitz, T. & Yellen, G. Two functionally distinct subsites for the binding of internal blockers to the pore of voltage-activated K⁺ channels. *Proc. Natl Acad. Sci. USA* **93**, 13357–13361 (1996).
- Holmgren, M., Jurman, M. E. & Yellen, G. Structure and function of the S4–S5 loop of the Shaker K⁺ channel examined through cysteine mutagenesis and chemical modification. *J. Gen. Physiol.* **108**, 195–206 (1996).
- Choi, K. L., Aldrich, R. W. & Yellen, G. Tetraethylammonium blockade distinguishes two inactivation mechanisms in voltage-activated K⁺ channels. *Proc. Natl Acad. Sci. USA* **88**, 5092–5095 (1991).
- Demo, S. D. & Yellen, G. The inactivation gate of the Shaker K⁺ channel behaves like an open-channel blocker. *Neuron* **7**, 743–753 (1991).
- MacArthur, M. W. & Thornton, J. M. Influence of proline residues on protein conformation. *J. Mol. Biol.* **218**, 397–412 (1991).
- Durell, S. R. & Guy, H. R. Atomic scale structure and functional models of voltage-gated potassium channels. *Biophys. J.* **62**, 238–247 (1992).
- Holmgren, M., Shin, K. S. & Yellen, G. The activation gate of a voltage-gated K⁺ channel can be trapped in the open state by an intersubunit metal bridge. *Neuron* **21**, 617–621 (1998).
- Gross, A. & MacKinnon, R. Agitoxin footprinting the Shaker potassium channel pore. *Neuron* **16**, 399–406 (1996).
- Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* **245**, 635–637 (1989).
- Perozo, E., Cortes, D. M., & Cuello, L. G. Structural rearrangements underlying K⁺ channel activation gating. *Science* **285**, 73–78 (1999).
- Kamb, A., Tseng-Crank, J. C. L. & Tanouye, M. A. Multiple products of the *Drosophila* Shaker gene may contribute to potassium channel diversity. *Neuron* **1**, 421–430 (1988).
- Hoshi, T., Zagotta, W. N. & Aldrich, R. W. Biophysical and molecular mechanisms of Shaker potassium channel inactivation. *Science* **250**, 533–538 (1990).
- López-Barneo, J., Hoshi, T., Heinemann, S. H. & Aldrich, R. W. Effects of external cations and mutations in the pore region on C-type inactivation of Shaker potassium channels. *Receptors Channels* **1**, 61–71 (1993).
- Jurman, M. E., Boland, L. M., Liu, Y. & Yellen, G. Visual identification of individual transfected cells for electrophysiology using antibody-coated beads. *BioTechniques* **17**, 876–881 (1994).
- Heginbotham, L. & MacKinnon, R. The aromatic binding site for tetraethylammonium ion on potassium channels. *Neuron* **8**, 483–491 (1992).
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Arch.* **391**, 85–100 (1981).
- Murrell-Lagnado, R. D. & Aldrich, R. W. Interactions of amino terminal domains of Shaker K channels with a pore blocking site studied with synthetic peptides. *J. Gen. Physiol.* **102**, 949–975 (1993).
- Armstrong, C. M. Time course of TEA⁺-induced anomalous rectification in squid giant axons. *J. Gen.*

- Physiol.* **50**, 491–503 (1966).
25. Swenson, R. P. Jr Inactivation of potassium current in squid axon by a variety of quaternary ammonium ions. *J. Gen. Physiol.* **77**, 255–271 (1981).
26. Gardner, S. & Thornton, J. I. I. protein structure database. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 1071–1077 (1998).
27. Antz, C. *et al.* NMR structure of inactivation gates from mammalian voltage-dependent potassium channels. *Nature* **385**, 272–275 (1997).

Supplementary Information is available from Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank Ki Soon Shin for performing some preliminary experiments with MTSES; M. MacArthur and J. Thornton for performing a database search; T. Ogren for help with the cells; B. Bean for comments on the manuscript; and members of the Yellen laboratory for helpful discussions. The work was supported by an NIH (NINDS) grant to G.Y., a fellowship from the Muscular Dystrophy Association to M.H. and a fellowship from the Spanish Ministry of Education to D.D.C.

Correspondence and requests for materials should be addressed to G.Y. (e-mail: gary_yellen@hms.harvard.edu).

KIR expression on self-reactive CD8⁺ T cells is controlled by T-cell receptor engagement

Bertrand Huard^{*†‡} & Lars Karlsson^{*}

^{*} The R.W. Johnson Pharmaceutical Research Institute, 3210 Merryfield Row, San Diego, California 92121, USA

[†] Laboratoire d'immunologie des tumeurs, Tour E2, 1er étage, Faculté de Pharmacie de Chatenay-Malabry, 5 rue Jean Baptiste Clément, 92296 Chatenay-Malabry, France

Natural killer cell tolerance is maintained by the interaction of killer inhibitory receptors (KIRs) with self-major histocompatibility complex class I gene products. A subset of T cells also expresses inhibitory receptors, but the functional significance of these receptors on T cells is unclear^{1–3}. Here we show that, in the absence of T-cell receptor (TCR) engagement, KIRs expressed on CD8⁺ T cells are slowly downregulated by KIR ligands expressed on antigen-presenting cells. The resulting expression levels of KIR are no longer able to inhibit T-cell function. In contrast, TCR engagement sustains KIR expression, and re-induces functional levels of KIR expression after ligand-induced downregulation of KIR. Our data indicate that KIR expression on CD8⁺ T cells *in vivo* may be maintained through continuous encounters with antigen. As KIR-mediated inhibition of T-cell activation can be bypassed at high antigen concentrations, dynamic KIR expression may mediate T-cell tolerance to self-antigens by sparing self-reactive T cells, thus enabling them to mediate potentially useful immune functions to quantitatively or qualitatively different antigens.

Human NK and T cells express two families of MHC class I reactive inhibitory receptors: immunoglobulin-like receptors (for example, KIR) directly interact with various MHC class I molecules (human leukocyte antigen (HLA)-A,-B,-C and -G)^{4–6}; and lectin-like receptors (CD94/NKG2), which interact with HLA-E-presenting signal-sequence peptides derived from other MHC class I molecules^{7–9}.

Although expression of CD94/NKG2 can be induced on T cells by certain cytokines¹⁰, expression of KIRs on natural killer (NK) and T cells is believed to be stable over time and independent of cellular activation^{10,11}. We observed that incubation of the KIR2D-expressing

CD8⁺ T-cell clone 1.15 (which is specific for an HLA-A2.1-restricted peptide derived from the melanocyte antigen tyrosinase¹²) with .221-A2.1-Cw7 cells (generated by transfection of the HLA-A, -B and -C-deficient cell line 721.221 (ref. 13) with HLA-A2.1 and HLA-Cw7 (a KIR2D ligand)) resulted in down-regulation of KIR2D from the T-cell surface when antigenic peptide was absent (Fig. 1a). After 24 hours, the KIR2D expression was reduced to 18% of the

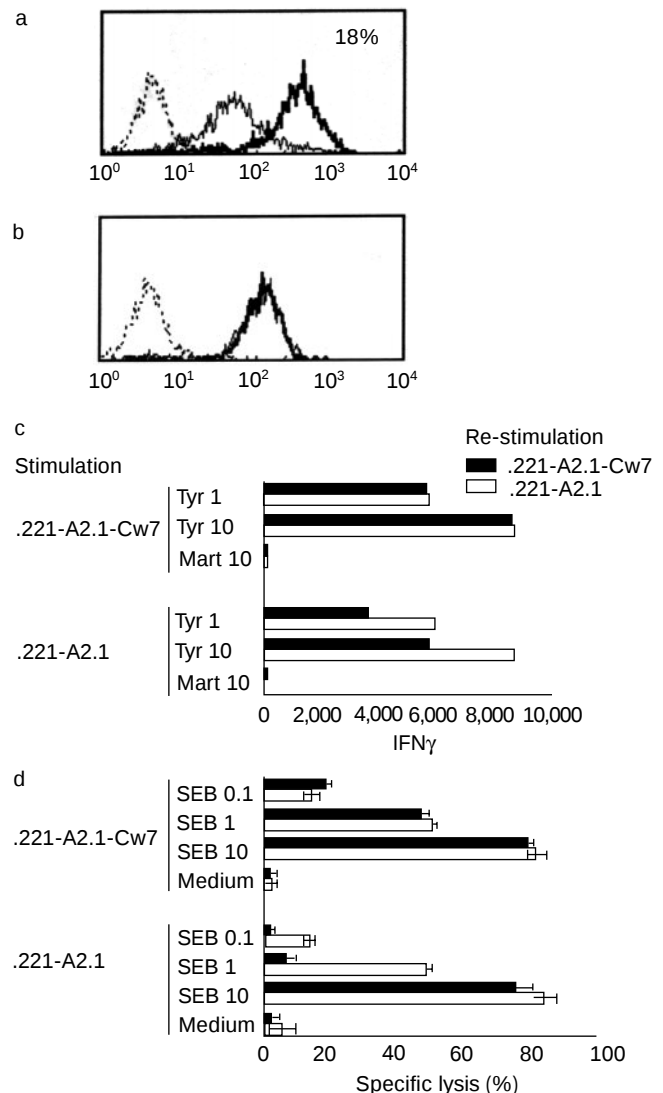


Figure 1 HLA ligands on APCs reduce KIR expression on T cells to non-functional levels. KIR2DL-expressing CD8⁺ T-cell clone 1.15 (**a**) or a KIR2DS-expressing CD8⁺ T-cell clone (**b**), both specific for a tyrosinase peptide presented by HLA-A2.1, were incubated with irradiated .221-A2.1 or .221-A2.1-Cw7 cells in the presence of an irrelevant peptide (mart-1). After 48 hours, KIR2DL expression was reduced to 18%, as assessed by immunofluorescence staining and flow cytometry analysis on gated T cells (BMA-031 staining). Thick and thin lines correspond to T cells incubated with .221-A2.1 and .221-A2.1-Cw7 cells, respectively. Dashed line represents background staining obtained with a control monoclonal antibody. **c**, KIR2D function after downregulation was assessed in a short-term re-stimulation assay. T-cell clone 1.15, previously incubated with the indicated APCs for 48 hours, was re-stimulated with .221-A2.1 (open bars) or with .221-A2.1-Cw7 (filled bars) cells loaded with peptides as indicated. Tyr, tyrosinase; Mart = mart-1, at 1 or 10 μ M. Intracytoplasmic IFN γ was stained, analysed by flow cytometry on gated T cells and quantified (arbitrary units). A representative experiment is shown. **d**, A SEB reactive KIR2D-expressing T-cell clone, B16, was incubated with .221-A2.1 or .221-A2.1-Cw7 cells as in **a**. After 48 hours, the cytotoxic activity of T cells against .221-A2.1 (open bars) or .221-A2.1-Cw7 (filled bars) cells pulsed with medium alone or pulsed with SEB (μ g ml⁻¹) was assessed in a ⁵¹Cr-release assay. Error bars, $\pm$ s.d.

[‡] Present address: Dermatology laboratory 5.222, Centre Médical Universitaire, 1 rue Michel Servet, CH-1211 Genève 4, Switzerland.

expression obtained after incubation with .221-A2.1 cells (721.221 cells expressing only HLA-A2.1). In contrast, CD2 and CD3 expression (monitored as controls) were unaltered, and the viability of the T cells (assessed by dye exclusion and annexin V staining) was not affected (data not shown). Similar KIR2D downregulation was observed for several other CD8⁺ T-cell lines and clones specific for the same tyrosinase peptide.

Sequence analysis showed that only long inhibitory forms of KIR2D (KIR2DL) were downregulated, and this phenomenon was not observed with a short form of KIR2D (KIR2DS) expressed by another tyrosine-peptide-specific CD8⁺ T-cell clone (Fig. 1b). After KIR downregulation, re-stimulation with the tyrosinase peptide resulted in similar interferon (IFN)- γ production whether the APCs expressed HLA-Cw7 or not (Fig. 1c, upper panel). In contrast, the IFN γ production by KIR2D-expressing T cells that had been pre-incubated with .221-A2.1 cells was inhibited by the presence of HLA-Cw7 on the antigen-presenting cells (APCs) during re-stimulation (Fig. 1c, lower panel).

Similar KIR downregulation (to 22% of the initial expression) was observed with the KIR2D-expressing T-cell clone B16 (of

unknown antigen specificity, but reactive with the staphylococcal enterotoxin B (SEB)), after incubation with HLA-Cw7-expressing APCs. No inhibitory effect of HLA-Cw7 on cytotoxic activity was detectable after .221-A2.1-Cw7-induced KIR2D downregulation. In contrast, the cytotoxicity of this clone was inhibited at lower concentrations of SEB after pre-incubation with .221-A2.1 cells (similar to other systems^{14,15}) (Fig. 1d). The inhibition mediated by HLA-Cw7 was not seen with KIR2D-negative control T cells (data not shown) and could be reversed by the addition of 4E, an HLA-C-reactive monoclonal antibody that blocks KIR–HLA-C interactions¹⁶. There could not be a role for HLA-E expressed on the surface of the .221-transfectants as the T-cell clones 1.15 and B16 did not express CD94/NKG2, and the blocking monoclonal antibody 4E does not react with HLA-E (see Supplementary Information).

Analysis by flow cytometry showed that both cell-surface and total KIR2D staining (after cell permeabilization) were decreased in T cells incubated for 24 hours with .221-A2.1-Cw7 cells (Fig. 2a), indicating that the KIR2D did not accumulate intracellularly. In contrast, cells treated with bafilomycin A1 (which inhibits intracellular degradation by increasing lysosomal pH (ref. 17)) during the incubation with the APCs had increased total KIR2D staining, whereas the surface staining remained low, indicating that the internalized KIR may be accumulating intracellularly. Analysis by confocal microscopy showed that the cell-surface staining of KIR2D was markedly reduced in the presence of HLA-Cw7 and transformed into a ‘patchy’ pattern reminiscent of antibody-induced capping (compare upper and lower left panels, Fig. 2b). Bafilomycin A1 treatment increased the intracellular staining and some of the KIR2D colocalized with the lysosomal marker Lamp-1 (ref. 18) (Fig. 2b, lower right). The partial co-staining with Lamp-1 was expected, as bafilomycin also inhibits the transfer of internalized proteins to lysosomes¹⁹. These experiments show that the incubation of KIR2D-expressing T cells with HLA-Cw7-expressing APCs results in internalization and degradation of KIRs to a level at which no inhibitory function on T-cell reactivity can be detected.

The HLA-Cw7-induced downregulation of KIR2D was observed in the absence of TCR stimulation. The presence of antigen (and therefore the presence of TCR stimulation) changed the fate of KIR2D after interaction with HLA-Cw7-expressing cells. Thus, loading of HLA-A2.1 with tyrosine peptide also resulted in maintained high KIR2D expression on the surface of T-cell clone 1.15 in the presence of HLA-Cw7 (Fig. 3a). The pronounced KIR2D downregulation in the presence of an irrelevant peptide (mart-1) was confirmed. The kinetics of the KIR2D downregulation was slow; it was barely detectable after 5 hours and peaked between 24 and 48 hours. For comparison, TCR downregulation on T-cell clone 1.15 was maximal after 5 hours (data not shown), which is consistent with published data²⁰. Incubation of .221-A2.1-Cw7 cells with 50 $\mu\text{g ml}^{-1}$ of SEB also resulted in maintained high KIR2D expression on T-cell clone B16, whereas incubation with 5 and 0.5 $\mu\text{g ml}^{-1}$ of SEB resulted in a substantial inhibition of KIR2D downregulation compared with the cells incubated with .221-A2.1-Cw7 in the absence of SEB (Fig. 3b).

Re-induction of KIR expression after ligand-induced downregulation was analysed after pre-incubation for 48 hours in the presence of .221-A2.1 or .221-A2.1-Cw7. Figure 4a shows the kinetics of KIR2D re-expression for T-cell clone B16. In the presence (but not in the absence) of SEB, KIR2D was re-expressed on the T cells that had been pre-incubated with .221-A2.1-Cw7, whereas KIR expression was further increased on the cells pre-incubated with .221-A2.1. The KIR2D surface re-expression in the presence of SEB was clearly detectable after 24 hours (but not after 6 hours) and a complete re-induction was achieved after 48 hours. When KIR2D was fully re-induced, we observed a potent inhibition of the cytotoxic activity in the presence of HLA-Cw7 on the target cells (Fig. 4b, compare Fig. 1c). Similar KIR re-expression was also seen

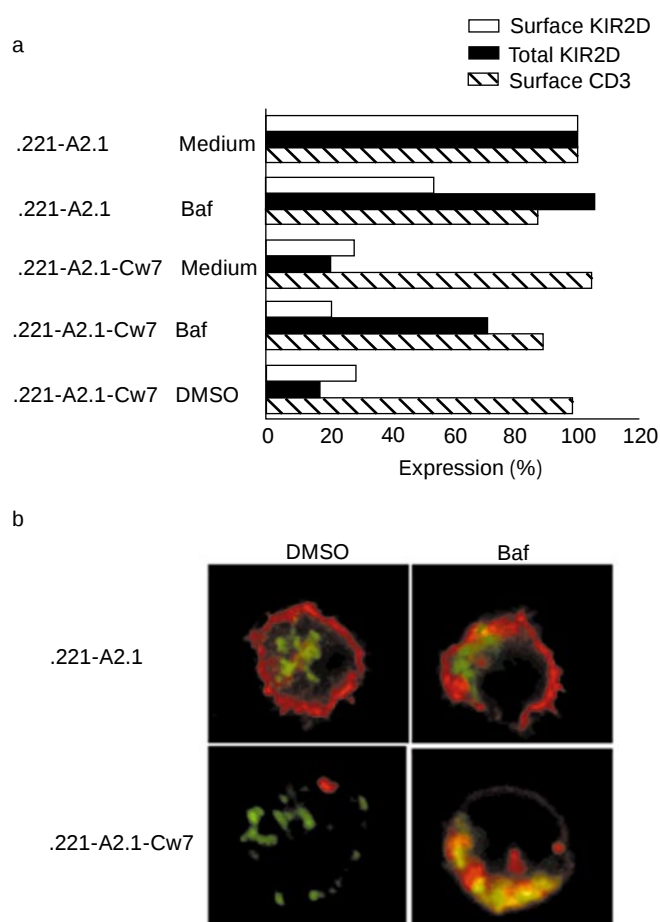


Figure 2 KIR2D is degraded after HLA-Cw7-induced downregulation. KIR2D⁺ T cells were incubated with bafilomycin A1 (1 μM), solvent alone (DMSO) or medium alone for 24 hours with the indicated APCs. **a**, Cell-surface KIR2D expression on non-permeabilized cells and total (cell-surface and cytoplasmic) KIR2D expression on permeabilized cells were assessed by immunostaining and flow cytometry analysis. The percentage of remaining surface (open bars) and total (filled bars) KIR2D expression is shown. One hundred per cent expression was defined arbitrarily as the mean fluorescence obtained with T cells incubated in medium with .221-A2.1 cells. Control CD3 surface expression is shown as hatched bars. **b**, Confocal images of T cells stained with monoclonal antibody GL183 (anti-KIR2D)(red) and rabbit anti-Lamp-1 (green) after incubation in bafilomycin A1 (1 μM) or, as a control, medium with solvent (DMSO). Baf, bafilomycin A1.

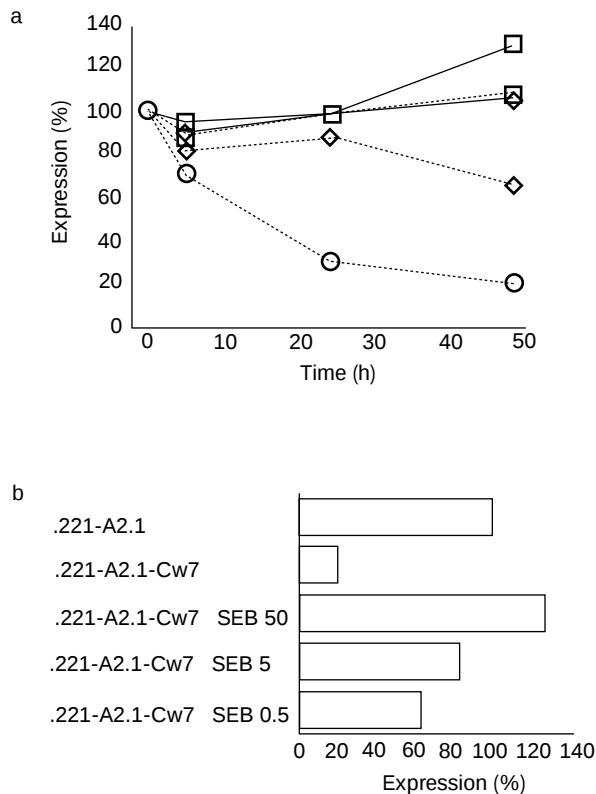


Figure 3 TCR triggering sustains KIR expression on T cells in the presence of HLA ligands on APCs. **a**, T-cell clone 1.15 was stimulated with .221-A2.1 (solid lines) or with .221-A2.1-Cw7 (dashed lines) cells loaded with 10 μ M of tyrosinase peptide (squares), 1 μ M of tyrosinase peptide (diamonds) or 10 μ M of mart-1 peptide (circles). KIR2D expression level was assessed as in Fig. 1a. The percentage of remaining KIR2D on T cells after 5, 24 and 48 hours is shown. 100% expression was defined arbitrarily by the mean fluorescence obtained with T cells incubated with .221-A2.1 loaded with 10 μ M of mart-1 peptide. **b**, T-cell clone B16 was incubated with unpulsed .221-A2.1, unpulsed .221-A2.1-Cw7 or with .221-A2.1-Cw7 cells pulsed with different concentrations of SEB (in μ g ml⁻¹) for 48 hours. KIR2D expression level was assessed and quantified as in **a**.

using clone 1.15 after incubation with tyrosinase peptide (data not shown). These experiments show that HLA-Cw7 is able to induce lasting KIR2D downregulation from the T-cell surface, but that TCR engagement can counteract KIR downregulation and cause the re-induction of functional levels of KIR2D expression. Considering the intracellular KIR degradation after HLA-Cw7 binding and the kinetics of KIR re-induction, there is probably a requirement for *de novo* KIR synthesis.

How KIR expression is initially induced remains unknown, but KIR⁺ T cells in blood are mostly CD8⁺ cells with an activated or memory phenotype^{2,21}. Thus, it is possible that a fraction of cells expresses KIRs during activation, rendering them less susceptible to exhaustion by overstimulation, and therefore giving them a selective advantage during the phase of memory-cell development. Once established, KIR expression appears to be dependent both on the presence of KIR ligands on APCs and on triggering through the TCR. This finding is consistent with both MHC class-I-induced modulation of inhibitory receptor expression on mouse NK cells²² and the increase in cell surface expression of CD94/NKG2 molecules after activation of $\gamma\delta$ T cells²³. In addition, KIR expression on dull KIR⁺ peripheral T cells²⁴ can be upregulated by lectin-driven T-cell activation (our unpublished data). The requirement for TCR stimulation suggests that peripheral blood T cells expressing KIRs may represent cells that are continuously exposed to antigens. Whether these antigens are persistent foreign antigens or self-antigens is unknown, but KIR2D-expressing CD8⁺ T cells specific for self-antigens can be isolated from healthy donors (ref. 21; and

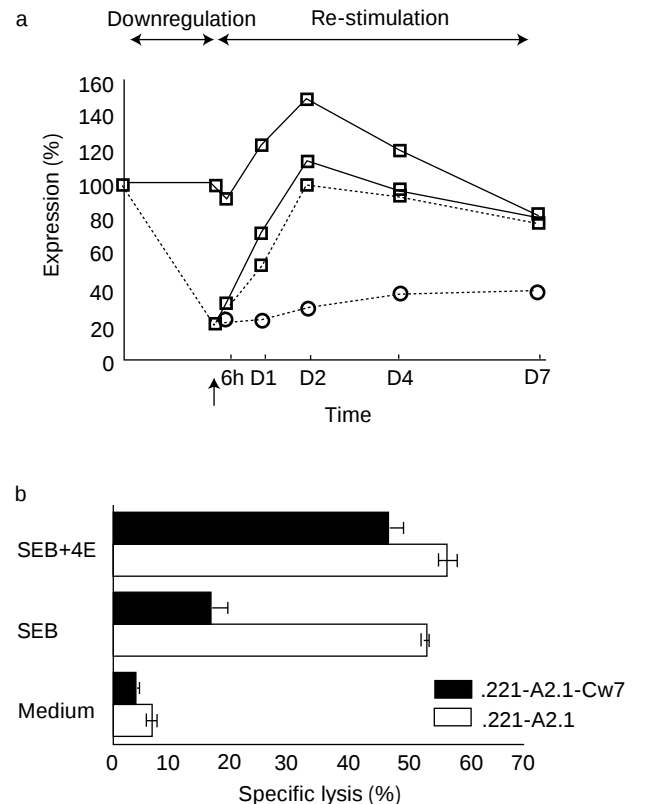


Figure 4 Re-induction of KIR expression after downregulation upon TCR triggering.

a, T-cell clone B16 was incubated with irradiated .221-A2.1 (solid lines) or .221-A2.1-Cw7 (dashed lines) for 48 hours ('Downregulation'). After this time (arrow), T cells were re-stimulated ('Re-stimulation') by addition of irradiated .221-A2.1-Cw7 (circles, dashed lines), .221-A2.1-Cw7 pulsed with SEB at 15 μ g ml⁻¹ (squares, dashed lines) or .221-A2.1 pulsed with SEB at 15 μ g ml⁻¹ (squares, solid lines). The percentage of KIR2D expression on T cells at the indicated times is shown. T cells incubated with .221-A2.1 and re-stimulated with unpulsed .221-A2.1 were used to arbitrarily define 100% KIR2D expression in this experiment. **b**, KIR function after re-induction was assessed in a short-term re-stimulation assay. B16 T cells incubated with .221-A2.1-Cw7 cells for 48 hours and re-stimulated with SEB-pulsed .221-A2.1-Cw7 cells for 7 days were used as effector cells in a ⁵¹Cr-release assay. Unpulsed (medium) or SEB-pulsed (1 μ g ml⁻¹) .221-A2.1 (open bars) and .221-A2.1-Cw7 (filled bars) cells were used as target cells. The HLA-C-reactive monoclonal antibody 4E was used to block KIR2D-HLA-Cw7 interactions. Error bars, $\pm$ s.d.

B.H. and L.K., manuscript submitted). Thus, KIR expression may serve as a mechanism to control the activity of self-cross-reactive T cells in order to spare potentially useful T-cell specificities that can be activated by foreign antigens with higher affinity for the TCR, or that are presented at higher concentrations. KIR expression may therefore represent an active silencing process inducing not only NK cell tolerance, but also T-cell tolerance to self-antigens accessible in the periphery. □

Methods

Cell lines and cloning

Purified peripheral CD8⁺ T cells from HLA-A2.1-expressing healthy donors were primed *in vitro* with HLA-A2.1, CD54 and CD80 transfected *Drosophila melanogaster* S2 cells²⁵. S2 cells were loaded with 100 μ M of tyrosinase peptide in the presence of 5 μ g ml⁻¹ of human β 2 microglobulin before T-cell stimulation. Lines were expanded by repeated stimulations with autologous irradiated adherent mononuclear cells loaded with 20 μ M tyrosinase peptide and human β 2 microglobulin. Interleukin (IL)-2 at 10 U ml⁻¹ and IL-7 at 30 U ml⁻¹ were added to the culture twice a week. The GL183 monoclonal antibody (Coulter) was used to purify T cells expressing KIR2D2 and KIR2D3 (both reactive with HLA-Cw7) by flow cytometry (KIR2D⁺ fraction). The KIR2D⁺ T cells were further expanded and cloned with irradiated .221-A2 cells loaded with 10 μ M of the relevant peptide. Peptides used were tyrosinase (AA 368-376) and control mart-1 (AA 27-35). The

KIR2D-expressing T-cell clone B16, which is reactive with SEB, was selected from a panel of KIR2D-expressing T-cell clones. These T-cell clones were generated after purification by flow cytometry of KIR2D⁺ peripheral T cells, followed by expansion using PHA (phytohaemagglutinin) and IL-2. SEB activity was assessed in a cytotoxicity assay using recombinant SEB (Toxin Technology). CD94/NKG2 expression was assessed with monoclonal antibody Hp-3D9 (Ancell).

Oligonucleotide primers specific for conserved sequences in KIR2D were used to amplify the transmembrane and cytoplasmic domains. Forward and reverse primers were 5'-ACCTACAGATGCTTCGG-3' and 5'-GAACCTCCAAATGCTGAG-3', respectively. Complementary DNA from KIR2D⁺-sorted cytotoxic T lymphocyte (CTL) lines were used as template. Polymerase chain reaction products were cloned into pCR 2.1-TOPO (Invitrogen) by TA cloning and subjected to DNA sequencing.

Cell line .221-A2.1 was supertransfected with a pHebo vector containing genomic HLA-Cw*0702 by electroporation (250V, 500 µF, infinite shunt resistor) with a Biorad Gene Pulser. HLA-Cw7 expression was monitored with monoclonal antibody 4E.

Functional assays

Cytotoxicity assays were performed as described²⁶. For blocking experiments, monoclonal antibody 4E (HLA-C reactive) was used at 1/1,000 dilution of ascites. For intracellular IFN γ staining, CTLs were incubated at a concentration of 1×10^6 cells ml⁻¹ in the presence of $1 \mu\text{M}$ monensin with transfected .221 cells loaded with peptides. The E/T ratio was 2/1. After 5 h at 37 °C, the cells were washed once in PBS and surface stained with phycoerythrin/cyanin (PC5)-conjugated anti-TCR (BMA 031) and phycoerythrin (PE)-conjugated anti-KIRs. After washing, cells were subjected to intracellular cytokine staining using Cytofix/Cytoperm kit (Pharmingen) with the FITC-conjugated mouse anti-human IFN γ (clone B27). We quantified the IFN γ produced in the assay as arbitrary units of IFN γ produced per 100 cells using (number IFN γ ⁺ cells $\times$ mean fluorescence of IFN γ ⁺ cells/number total cells analysed) $\times$ 100. A validation of this quantification was performed in peptide titration and inhibition experiments (data not shown).

Immunostaining and microscopy

For intracellular KIR2D staining, T cells were pretreated with bafilomycin A1 ($1 \mu\text{M}$ in DMSO; Sigma) for 30 min then incubated with transfected .221 cells. After 24 hours, cells were permeabilized in PBS with 1% saponin, immunostained with GL183 monoclonal antibody and analysed by flow cytometry. For confocal imaging, cells were attached to poly-L-lysine coated slides, fixed with 3% formaldehyde-PBS and washed with 50 mM NH₄Cl-PBS. Antibody incubations were made in PBS with 0.6% fish skin gelatin and 0.2% saponin for permeabilization. FITC-labelled anti-rabbit immunoglobulin (Cappel) and Texas-Red labelled anti-mouse (Molecular Probes) secondary reagents were used. Fluorescent cells were imaged using an Olympus confocal microscope.

Received 19 August; accepted 3 November 1999.

- Phillips, J. H., Gumperz, J. E., Parham, P. & Lanier, L. L. Superantigen-dependent, cell-mediated cytotoxicity inhibited by MHC class I receptors on T lymphocytes. *Science* **268**, 403–405 (1995).
- Mingari, M. C. *et al.* Human CD8⁺ T lymphocyte subsets that express HLA class I-specific inhibitory receptors represent oligoclonally or monoclonally expanded cell populations. *Proc. Natl Acad. Sci. USA* **93**, 12433–12438 (1996).
- Mingari, M. C. *et al.* HLA-class I-specific inhibitory receptors in human cytolytic T lymphocytes: molecular characterization, distribution in lymphoid tissues and co-expression by individual T cells. *Int. Immunol.* **9**, 485–491 (1997).
- Lanier, L. L. NK cell receptors. *Annu. Rev. Immunol.* **16**, 359–393 (1998).
- Rajagopalan, S. & Long, E. O. A human histocompatibility leukocyte antigen (HLA)-G-specific receptor expressed on all natural killer cells. *J. Exp. Med.* **189**, 1093–1100 (1999).
- Ponte, M. *et al.* Inhibitory receptors sensing HLA-G1 molecules in pregnancy: decidua-associated natural killer cells express LIR-1 and CD94/NKG2A and acquire p49, an HLA-G1-specific receptor. *Proc. Natl Acad. Sci. USA* **96**, 5674–5679 (1999).
- Lee, N. *et al.* HLA-E is a major ligand for the natural killer inhibitory receptor CD94/NKG2A. *Proc. Natl Acad. Sci. USA* **95**, 5199–5204 (1998).
- Braud, V. M. *et al.* HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. *Nature* **391**, 795–799 (1998).
- Borrego, E., Ulbrecht, M., Weiss, E. H., Coligan, J. E. & Brooks, A. G. Recognition of human histocompatibility leukocyte antigen (HLA)-E complexed with HLA class I signal sequence-derived peptides by CD94/NKG2 confers protection from natural killer cell-mediated lysis. *J. Exp. Med.* **187**, 813–818 (1998).
- Mingari, M. C., Moretta, A. & Moretta, L. Regulation of KIR expression in human T cells: a safety mechanism that may impair protective T-cell responses. *Immunol. Today* **19**, 153–157 (1998).
- Moretta, A. *et al.* Receptors for HLA class-I molecules in human natural killer cells. *Annu. Rev. Immunol.* **14**, 619–648 (1996).
- Brichard, V. *et al.* The tyrosinase gene codes for an antigen recognized by autologous cytolytic T lymphocytes on HLA-A2 melanomas. *J. Exp. Med.* **178**, 489–495 (1993).
- Shimizu, Y. & DeMars, R. Production of human cells expressing individual transferred HLA-A, -B, -C genes using an HLA-A, -B, -C null human cell line. *J. Immunol.* **142**, 3320–3328 (1989).
- Carena, I., Shamshiev, A., Donada, A., Colonna, M. & De Libero, G. Major histocompatibility complex class I molecules modulate activation threshold and early signaling of T cell antigen receptor- $\gamma\delta$ stimulated by nonpeptidic ligands. *J. Exp. Med.* **186**, 1769–1774 (1997).
- Le Dren, E. *et al.* Inhibition of antigen-induced T cell response and antibody-induced NK cell cytotoxicity by NKG2A. Association of NKG2A with SHP-1 and SHP-2 protein-tyrosine phosphatases. *Eur. J. Immunol.* **28**, 264–276 (1998).
- Falk, C. S., Schendel, D. J. Expression of HLA-C molecules confers target cell resistance to some non-major histocompatibility complex-restricted T cells in a manner analogous to allospecific natural killer cells. *J. Exp. Med.* **182**, 1005–1018 (1995).
- Valitutti, S., Muller, S., Salio, M. & Lanzavecchia, A. Degradation of T cell receptor (TCR)–CD3-zeta

complexes after antigenic stimulation. *J. Exp. Med.* **185**, 1859–1964 (1997).

- Fukuda, M., Viitala, J., Matteson, J. & Carlsson, S. R. Cloning of cDNAs encoding human lysosomal membrane glycoproteins, h-lamp-1 and h-lamp-2. Comparison of their deduced amino acid sequences. *J. Biol. Chem.* **263**, 18920–18928 (1988).
- van Deurs, B., Holm, P. K. & Sandvig, K. Inhibition of the vacuolar H⁺-ATPase with bafilomycin reduces delivery of internalized molecules from mature multivesicular endosomes to lysosomes in HEP-2 cells. *Eur. J. Cell Biol.* **69**, 343–350 (1996).
- Valitutti, S., Muller, S., Cella, M., Padovan, E. & Lanzavecchia, A. Serial triggering of many T-cell receptors by a few peptide–MHC complexes. *Nature* **375**, 148–151 (1995).
- Speiser, D. E. *et al.* CD28-negative cytolytic effector T cells frequently express NK receptors and are present at variable proportions in circulating lymphocytes from healthy donors and melanoma patients. *Eur. J. Immunol.* **29**, 1990–1999 (1999).
- Höglund, P. *et al.* Host MHC class I gene control of NK-cell specificity in the mouse. *Immunol. Rev.* **155**, 11–28 (1997).
- Boullier, S. *et al.* Phosphoantigen activation induces surface translocation of intracellular CD94/NKG2A class I receptor on CD94-peripheral V γ V δ T cells but not on CD94⁺ thymic or mature gamma delta T cell clones. *Eur. J. Immunol.* **28**, 3399–3410 (1998).
- Gumperz, J. E., Valiante, N. M., Parham, P., Lanier, L. L. & Tyan, D. Heterogenous phenotypes of expression of the NKB1 natural killer cell class I receptor among individuals of different human histocompatibility leukocyte antigens types appear genetically regulated, but not linked to major histocompatibility complex haplotype. *J. Exp. Med.* **183**, 1817–1827 (1996).
- Cai, Z. *et al.* Transfected *Drosophila* cells as a probe for defining the minimal requirements for stimulating unprimed CD8⁺ T cells. *Proc. Natl Acad. Sci. USA* **93**, 14736–14741 (1996).
- Huard, B., Prigent, P., Tournier, M., Bruniquel, D. & Triebel, F. CD4/major histocompatibility complex class II interaction analyzed with CD4⁺ and lymphocyte activation gene-3 (LAG-3)-Ig fusion proteins. *Eur. J. Immunol.* **25**, 2718–2721 (1995).

Acknowledgements

We thank A. Vitiello for 721.221 transfected with HLA A2.1; E. Noessner for the HLA-Cw*0702 construct; Soo Young Yang for the mAb 4E; and S. Carlsson for the anti-lamp-1 antiserum. We thank E. Long for his critical reading of the manuscript. B. Huard thanks F. Triebel for hosting the final experiments of this project in his laboratory.

Correspondence and requests for materials should be addressed to B.H. (e-mail: huard@cmu.unige.ch).

Regulation of carbamoyl phosphate synthetase by MAP kinase

Lee M. Graves*†, Hedeel I. Guy‡, Piotr Kozlowski*, Min Huang*, Eduardo Lazarowski*, R. Marshall Pope§, Matthew A. Collins*, Erik N. Dahlstrand*, H. Shelton Earp III*† & David R. Evans‡

* Departments of Pharmacology and § Psychiatry and the † Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7365, USA

‡ Department of Biochemistry and Molecular Biology, Wayne State University, 936 Mary Ellen Jones Building, Detroit, Michigan 48201, USA

The *de novo* synthesis of pyrimidine nucleotides is required for mammalian cells to proliferate. The rate-limiting step in this pathway is catalysed by carbamoyl phosphate synthetase (CPS II), part of the multifunctional enzyme CAD^{1,2}. Here we describe the regulation of CAD by the mitogen-activated protein (MAP) kinase cascade. When phosphorylated by MAP kinase *in vitro* or activated by epidermal growth factor *in vivo*, CAD lost its feedback inhibition (which is dependent on uridine triphosphate) and became more sensitive to activation (which depends upon phosphoribosyl pyrophosphate). Both these allosteric regulatory changes favour biosynthesis of pyrimidines for growth². They were accompanied by increased epidermal growth factor-dependent phosphorylation of CAD *in vivo* and were prevented by inhibition of MAP kinase. Mutation of a consensus MAP kinase phosphorylation site abolished the changes in CAD allosteric regulation that were stimulated by growth factors. Finally, consistent with an effect of MAP kinase signalling on CPS II activity, epidermal growth factor increased cellular uridine triphosphate and this increase was reversed by inhibition of MAP kinase. Hence

these studies may indicate a direct link between activation of the MAP kinase cascade and *de novo* biosynthesis of pyrimidine nucleotides.

In mammalian cells, the first, second and third steps in the *de novo* synthesis of uridine nucleotides are catalysed by a single protein, CAD¹, which contains three enzymatic activities: carbamoyl phosphate synthetase, aspartate transcarbamoylase and dihydroorotase. The rate-limiting reaction is catalysed by carbamoyl phosphate synthetase, which is also the site of allosteric regulation by uridine 5'-triphosphate (UTP) and phosphoribosyl pyrophosphate (PRPP)². CPS II is related to the mitochondrial enzyme carbamoyl phosphate synthetase (CPS I), a component of the urea metabolic cycle³. The biosynthesis of pyrimidines is essential for cell proliferation, and CPS II activity is correlated with increased growth in normal and tumour cell lines^{2,4}. Total CAD activity can be increased as a result of transcription or gene amplification⁵; however, the regulation of CAD (CPS II) by acute growth signals has not been investigated.

The MAP kinase (MAPK) cascade is a ubiquitous protein phosphorylation pathway that is required for cell proliferation⁶ and is activated by many growth-promoting stimuli. MAPK phosphorylates both nuclear (for example, Myc, Elk-1) and non-nuclear

proteins (for example, cytosolic phospholipase A2)^{7,8}. While sequencing proteins that were phosphorylated by MAPK *in vitro*, we identified one of these as the mitochondrial enzyme carbamoyl phosphate synthetase (CPS I). MAPK preferentially phosphorylates substrates in a proline-directed manner. Inspection of peptide sequences from CPS I and CPS II (CAD) revealed that a potential consensus MAPK phosphorylation site (PITP)⁹ was conserved at amino acids 488 and 456, respectively (Fig. 1a). Incubation of CAD with active MAPK (Erk 2) and γ -³²P ATP showed that CAD was phosphorylated to a stoichiometry of about 1.2 mole Pi per mole CAD *in vitro* (Fig. 1b). Complete amino-acid hydrolysis and two-dimensional phosphoamino-acid mapping¹⁰ of the MAPK-phosphorylated CAD showed that phosphorylation occurred primarily on threonine *in vitro* (Fig. 1c). Digestion of the phosphorylated CAD protein by trypsin, followed by nano-electrospray mass spectrometry and two sequential stages of mass selection (MS/MS) scans for combined loss of water and phosphite¹¹, identified a doubly charged peptide with the expected mass (*m/z* 1,014.7) of residues 450–465 incorporating a single phosphoryl group (R.M.P., unpublished data; see Supplementary Information).

We investigated the regulation of CAD in response to MAPK activation *in vivo*. CAD is allosterically inhibited by the product of the pyrimidine pathway, UTP¹². We incubated rat liver epithelial cells (GN4) with epidermal growth factor (EGF) to activate MAPK, and measured the glutamine-dependent CPS activity of CAD as a function of UTP concentration¹³. Adding UTP to lysates from serum-starved cells inhibited the CPS activity by around 40–50% (Fig. 2a). The CPS activity from EGF-treated cells was not inhibited by UTP; instead, UTP significantly increased the CPS activity at

a

Gene	Accession no.	Sequence
Human CAD	D78586	ADKVYFLPITPHYVTQVI
Hamster CAD	J05503	ADKVYFLPITPHYVTQVI
<i>Squalus acanthias</i> CAD	U18868	ADKVYFLPITPEYVTQVI
<i>Drosophila melanogaster</i> CAD	X04813	ADKCYFLPLPHYVEQVI
<i>Saccharomyces cerevisiae</i> URA2	M27174	ADKVYFVPVIAEFVRKVI
Human CPS I	D90282	ADTVYFLPITPQFVTEVI
Rat CPS I	M11710	ADAVYFLPITPQFVTEVI
<i>Rana catesbeiana</i> CPS I	U05193	ADTVYFLPITPQFVTEVI
<i>Saccharomyces cerevisiae</i> CPA2	K01178	ADKIYLPVITPEYITYII

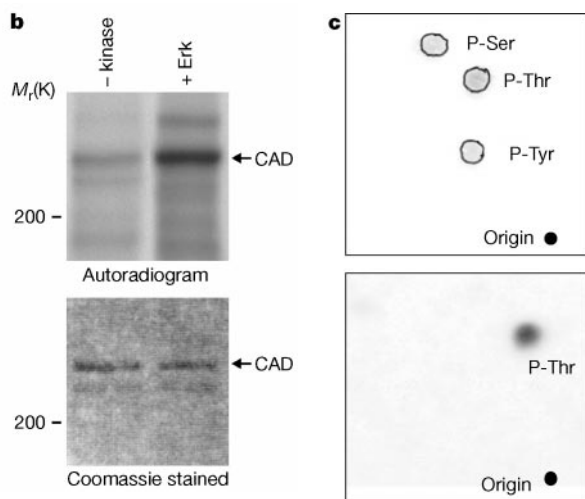


Figure 1 MAPK phosphorylates CAD *in vitro*. **a**, The predicted amino-acid sequences from CPS II (CAD) and CPS I were analysed for MAPK phosphorylation sites (the GenBank accession numbers corresponding to the gene products are listed). A threonine residue (underlined) was conserved within a sequence of amino acids that represents a consensus MAPK phosphorylation site (bold residues). **b**, CAD was immunoprecipitated from serum-starved GN4 cells and incubated with active Erk 2 and γ -³²P ATP (see Methods). The phosphorylated proteins were separated on SDS-PAGE (8% acrylamide), and the amount of radioactivity incorporated into CAD was determined by autoradiography and scintillation counting of the excised CAD bands. The amount of CAD was estimated by staining with Coomassie blue. **c**, CAD was phosphorylated with Erk 2 as in **b** and phosphoamino-acid analysis performed as described¹⁰. The phosphoamino-acid standards (P-Thr, P-Ser, P-Tyr) were visualized by ninhydrin (upper panel), and the phosphorylated amino acids were detected by autoradiography (lower panel).

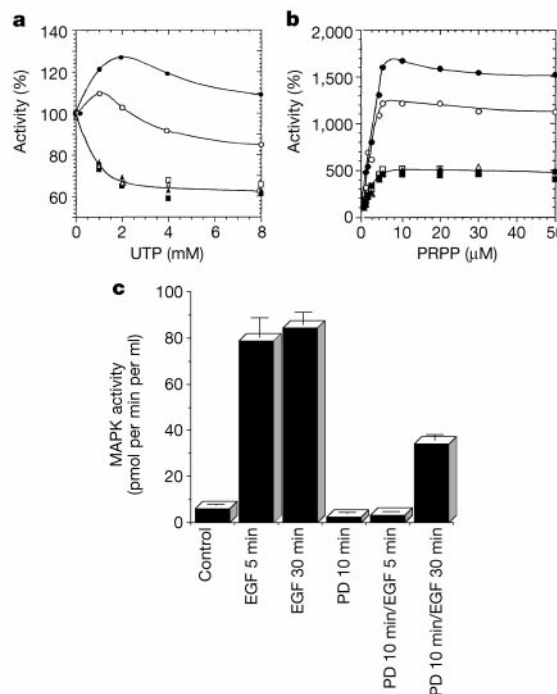


Figure 2 EGF regulates the allosteric properties of CAD. The glutamine-dependent CPSase activity of CAD was measured as a function of UTP (**a**) or PRPP (**b**) concentration. Concentration in lysates from serum-starved GN4 cells (filled squares), cells treated with 50 μ M PD98059 (open squares), cells incubated with EGF (40 ng ml⁻¹) for 5 min in the absence (open circles) or presence (filled circles) of the inhibitor. Activity is plotted as percentage of total activity obtained without added ligand. Absolute activities of CAD were 0.046 ± 0.0021 nmol min⁻¹. **c**, MAPK activity was measured in immunoprecipitates from serum-starved, EGF-treated, PD98059 (PD; 50 μ M, 10 min) treated or PD98059 then EGF-treated GN4 cells (see Methods). MAPK activity is plotted as pmol per min per ml of resuspended immunoprecipitate (*n*=2, mean $\pm$ s.e.).

some concentrations (2 mM). This effect was enhanced with increased exposure to EGF and occurred in parallel with activation of MAPK (Fig. 2c). Moreover, incubating cells with the MAPK kinase inhibitor PD98059 (ref. 14) prevented the EGF-dependent activation of MAPK and restored the UTP-dependent inhibition of CPS II (Fig. 2a).

PRPP allosterically activates CAD by interacting with the CPS II domain¹². Adding PRPP to lysates of serum-starved cells increased the glutamine-dependent CPS activity of CAD by about 400–500% (Fig. 2b). In comparison, PRPP increased the CPS II activity from EGF-treated cells to a much greater extent (up to 1,500%). This effect depended on the length of exposure to EGF and was reversed by incubating cells with PD98059 (Fig. 2b). Treatment with platelet-derived growth factor-BB (PDGF) also activated MAPK and enhanced the activation of CPS II by PRPP, showing that regulation of CAD was not restricted to EGF receptor signalling in these cells (data not shown).

We tested whether EGF could influence the phosphorylation of CAD *in vivo*. Immunoprecipitation of CAD from lysates of ³²P-orthophosphate-labelled cells showed that CAD was weakly phosphorylated in serum-starved cells (Fig. 3a). Epidermal growth factor stimulated a 2–3-fold increase in the phosphorylation of CAD that paralleled the changes in the allosteric properties of CAD (see Fig. 2a,b). The maximal increase in phosphorylation occurred after 30 min. Incubating cells with PD98059 prevented this increase, consistent with MAPK mediating the phosphorylation of CAD *in vivo* (Fig. 3a,b). To determine whether MAPK regulated the allosteric properties of CAD, we incubated purified CAD with active MAPK (Erk 2) and measured the allosteric regulation by UTP and PRPP. Phosphorylating CAD with MAPK *in vitro* reduced the inhibition of CPS II by UTP (to less than 10%; Fig. 4a) and increased activation by PRPP up to 3–4-fold compared with non-phosphorylated CAD (Fig. 4b).

As our studies indicated that Thr 456 in CAD was a potential site of phosphorylation, we used site-directed mutagenesis to replace Thr 456 with alanine. We transfected the wild-type and mutant CAD genes into a cell line deficient in CAD (G9c; ref. 15) and obtained G418-resistant stable cell lines (G9c-Thr 456 and G9c-Ala 456) expressing similar amounts of the wild-type and mutant CAD proteins and activity (Fig. 4c; and data not shown). Because G9c cells lack EGF receptors, the G9c-Thr 456 and G9c-Ala 456 cells were incubated with PDGF, resulting in equivalent activation of

MAPK in these cells (data not shown). In lysates from the PDGF-treated G9c-Thr 456 cells, PRPP increased the activation of CPS II by up to 1,800% over that found in untreated cells. In contrast, PDGF did not enhance the CPS II activity in lysates from the G9c-Ala 456 cells; this mutant was activated by PRPP to the same extent as the starved wild-type (G9c-Thr 456) cells (200–300%) (Fig. 4c). We also examined the UTP-dependent inhibition of CAD in these cells. PDGF treatment of the G9c-Thr 456 cells reversed inhibition by UTP (Fig. 4d), whereas PDGF had no effect on the UTP-dependent inhibition of CPS II from the G9c-Ala 456 cells.

CAD is the rate-limiting enzyme in the *de novo* synthesis of uridine nucleotides². Therefore we determined the effect of MAPK activation on intracellular (UTP) levels¹⁶. Treatment of serum-starved GN4 cells with EGF initially caused a slight decline in cellular UTP concentrations but stimulated a significant increase after 30–45 min (Fig. 5). This occurred slightly more slowly than the effects of EGF on the allosteric regulation of CAD (Fig. 2a,b); however, consistent with a role for MAPK in this process, incubation with PD98059 blocked the EGF-stimulated increase in UTP accumulation (Fig. 5). In comparison, intracellular ATP was not changed by treatment with EGF or PD98059 (data not shown).

We have described an unprecedented role for MAPK in the regulation of CAD and the essential process of *de novo* biosynthesis of pyrimidine nucleotides. Given the estimated intracellular concentrations of UTP (0.567 ± 0.460 mM) and PRPP (100 μM), the influence of MAPK signalling on the allosteric regulation of CAD and the *de novo* synthesis of pyrimidines is predicted to be significant^{2,17}. Our findings indicate that CAD may be phosphorylated and regulated by MAPK; two-dimensional phosphopeptide mapping experiments revealed common peptides from CAD phosphorylated *in vitro* or in CAD from EGF-treated cells

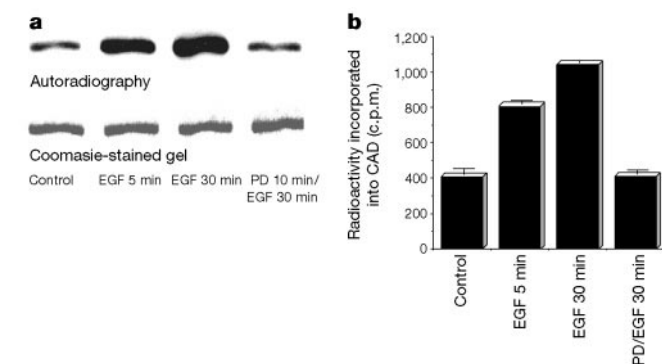


Figure 3 CAD is phosphorylated *in vivo* in response to EGF and MAPK activation. **a**, Serum-starved GN4 cells were labelled with ³²P-orthophosphate, stimulated with EGF for the duration indicated and immunoprecipitated from the cell lysates with anti-CAD antibodies. In some experiments, cells were incubated with PD98059 (50 μM, 10 min) before EGF was added. The immunoprecipitates were applied to SDS-PAGE and the phosphorylated CAD detected by autoradiography. A representative autoradiograph of four experiments is shown. **b**, The amount of radioactivity in CAD was determined by excision of phosphorylated protein and liquid scintillation counting. The data plotted represent the s.e. of duplicate experiments.

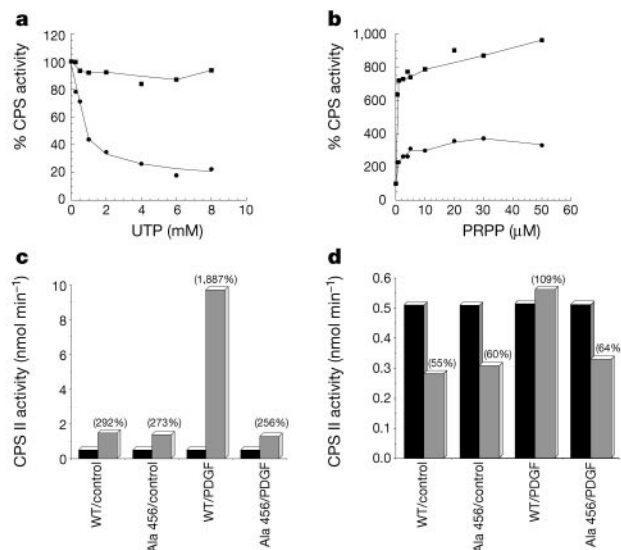


Figure 4 CAD is regulated by MAPK phosphorylation. CAD was purified from 165–23 cells, incubated with (squares) or without (circles) active Erk 2 and Mg-ATP (see Methods). The glutamine-dependent CPSase activity of CAD was measured as a function of UTP (**a**) or PRPP concentration (**b**) (ref. 13) and the activity plotted as percentage of the CPS activity without added ligand. The absolute activity of CAD incubated with or without MAPK was 0.034 and 0.039 nmol min⁻¹, respectively. **c**, **d**, Stable cell lines expressing wild-type (WT) (G9c-Thr 456) or mutant CAD (G9c-Ala 456) were serum-starved and treated with PDGF for the time indicated. Cell lysates were prepared and the glutamine-dependent CPSase activity of CAD was measured as a function of PRPP (**c**) or UTP (**d**) as described above. The activity is plotted as nmol min⁻¹, and the percentage of CPS II activity measured in the presence of PRPP or UTP is shown in parentheses. The activation of the wild-type CAD as a function of PRPP was 2–6-fold and was always abolished with the mutant.

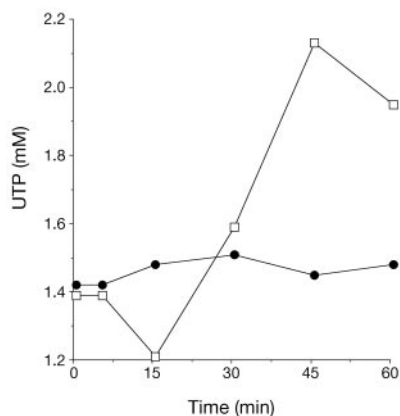


Figure 5 EGF stimulates UTP synthesis through a MAPK-dependent pathway. Serum-starved GN4 cells were incubated with EGF (40 ng ml⁻¹) with (circles) or without (squares) pre-treatment with PD98059 (50 μM) for the time indicated. The amount of UTP (mM) was determined by a coupled enzyme assay as described¹⁶ and calculated using an average cell volume for rat liver cells of 3.4 ml per 10⁹ cells (ref. 17). A representative of three trials is shown.

(P.K., unpublished data). Mutation of a potential MAPK phosphorylation site (Thr 456) abolished the growth-factor dependent regulation of CAD, further indicating that this may be an important regulatory site *in vivo*. However, it remains to be shown that Thr 456 is a major MAPK site of phosphorylation either *in vitro* or *in vivo*. CAD is phosphorylated by the cyclic AMP-dependent protein kinase (PKA), resulting in a loss of feedback inhibition by UTP and providing a mechanism by which cAMP may increase pyrimidine biosynthesis^{18–20}. However, in contrast to the enhanced PRPP activation that accompanies MAPK phosphorylation of CAD, PKA phosphorylation of a chimaeric CPSase suppresses the effect of PRPP²¹, potentially resulting in the inhibition of pyrimidine synthesis. Hence CAD may be a focal point for regulation by both these kinases. The contribution of these signals to the biosynthesis of pyrimidines merits further investigation. □

Methods

Cell growth, treatment and lysate preparation

GN4 rat liver epithelial cells (60-mm dishes) were grown, serum-starved and harvested in cell lysis buffer as described²². Cell lysates were clarified by centrifugation and assayed for protein content using Coomassie protein assay reagent (Pierce). CAD was immunoprecipitated from cell lysates using a rabbit polyclonal antibody and protein A agarose. The immunoprecipitates were washed twice in cell lysis buffer and once in PBS before phosphorylation or SDS–PAGE experiments.

In vitro phosphorylation of CAD with Erk 2 and *in vivo* labelling of CAD with ³²P-orthophosphate

For *in vitro* phosphorylation and CPS assays, CAD was either immunoprecipitated from cells or purified from a Syrian hamster overexpressing cell line (165-23) as described²³. CAD (4.8 μg) was phosphorylated with 20 ng purified, active recombinant Erk 2 (New England Biolabs) in a buffer containing 25 mM Tris, pH 7.5, 5 mM β-glycerophosphate, 0.1 mM Na₃VO₄, 2 mM DTT, 10 mM MgCl₂ and 200 μM ATP for 30 min at 30 °C. For phosphorylation stoichiometry experiments, CAD was incubated in a buffer containing 20 mM Hepes, pH 7.3, 10 mM β-glycerophosphate, 1.5 mM EGTA, 0.1 mM Na₃VO₄, 1 mM DTT, 10 mM MgCl₂ and 10 μM [γ-³²P ATP] (6 μCi per reaction) with or without 2 μg of active recombinant Erk 2 for 60 min at 30 °C. The radioactivity incorporated into CAD was quantified by excision of the radiolabelled proteins and liquid scintillation counting. Two-dimensional phosphoamino-acid analysis was performed as described¹⁰ after transfer of the phosphorylated CAD to Immobilon PVDF (Millipore).

For *in vivo* labelling experiments, 100-mm plates of serum-starved GN4 cells (90% confluent) were incubated with phosphate-free medium for 30 min, after which 0.5 mCi of ³²P-orthophosphate was added to each plate. The cells were allowed to incorporate ³²P-orthophosphate for 2 h and then treated with EGF or PD98059 as described above. The labelled cells were harvested in cell lysis buffer (0.7 ml) and immunoprecipitated as described above. After multiple washes in cell lysis buffer, the samples were applied to SDS–PAGE (8% acrylamide) and analysed by autoradiography.

Creation of stable cell lines expressing wild-type (Thr 456) and mutant (Ala 456) CAD proteins

We used PCR mutagenesis to convert Thr 456 to alanine with the primer CTACTTCCTTCCCATTCACCTCACTACGTAACCC and cloned the resulting DNA into pCDNA3.1 HisC (Invitrogen). Stable cell lines expressing the Thr 456 and Ala 456 CAD proteins were created by transfecting CAD-deficient cells (G9c; ref. 15)) with 5 μg DNA (pCDNA 3.1/HisC Thr 456 CAD or Ala 456 CAD) using Lipofectamine according to the manufacturer's procedure. Cells were grown in DMEM/F12 (Gibco-BRL) with 10% FBS, non-essential amino acids, PenStrep and 30 μM uridine. After two days the cells were split (1:10) and selected with G418 (0.5 mg ml⁻¹). Cells were selected with G418 for two weeks, after which uridine was no longer required for growth. After two additional passages, G418 was also removed from the medium.

Enzyme assays

We determined MAPK activity as described²² except that the assay was terminated by spotting the phosphorylated myelin basic protein on P-81 paper (Whatman), washed in 150 mM phosphoric acid (five changes), dried briefly and counted in a liquid scintillation counter. We assayed the glutamine-dependent CPS II activity and the allosteric regulation of CAD by UTP or PRPP using [¹⁴C NaHCO₃] as described¹³.

Measurement of UTP synthesis

GN4 liver cells were serum-starved, harvested in ice-cold 5% TCA and clarified by centrifugation. The supernatants were extracted with diethyl ether (three times) to remove TCA. The aqueous samples were neutralized and the amount of UTP in the aqueous phase determined using a coupled enzyme assay¹⁶.

Received 15 July; accepted 8 November 1999.

- Evans, D. R. in *Multidomain Proteins—Structure and Evolution* (eds Hardie, D. G. & Coggins, J. R.) 283–331 (Amsterdam, Elsevier, 1986).
- Jones, M. E. Pyrimidine nucleotide biosynthesis in animals: genes, enzymes, and regulation of UMP biosynthesis. *Annu. Rev. Biochem.* **49**, 253–279 (1980).
- Anderson, P. M. in *Nitrogen Metabolism and Excretion* (eds Walsh, P. J. & Wright, P. A.) 33–55 (CRC, Boca Raton, Florida, 1995).
- Aoki, T. & Weber, G. Carbamoyl phosphate synthetase (glutamine-hydrolyzing): increased activity in cancer cells. *Science* **212**, 463–465 (1981).
- Smith, K. A. *et al.* Regulation and mechanisms of gene amplification. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **347**, 49–56 (1995).
- Lewis, T. S., Shapiro, P. S. & Ahn, N. G. Signal transduction through MAP kinase cascades. *Adv. Cancer Res.* **74**, 49–139 (1998).
- Cobb, M. H. & Goldsmith, E. J. How MAP kinases are regulated. *J. Biol. Chem.* **270**, 14843–14846 (1995).
- Davis, R. J. Transcriptional regulation by MAP kinases. *Mol. Reprod. Dev.* **42**, 459–467 (1995).
- Gonzalez, F. A., Raden, D. L. & Davis, R. J. Identification of substrate recognition determinants for human ERK1 and ERK2 protein kinases. *J. Biol. Chem.* **266**, 22159–22163 (1991).
- van der Geer, P. & Hunter, T. Phosphopeptide mapping and phosphoamino acid analysis by electrophoresis and chromatography on thin-layer cellulose plates. *Electrophoresis* **15**, 544–554 (1994).
- Resing, K. A. & Ahn, N. G. Protein phosphorylation analysis by electrospray ionization–mass spectrometry. *Methods Enzymol.* **283**, 29–44 (1997).
- Shoaf, W. T. & Jones, M. E. Uridylic acid synthesis in Ehrlich ascites carcinoma. Properties, subcellular distribution, and nature of enzyme complexes of the six biosynthetic enzymes. *Biochemistry* **12**, 4039–4051 (1973).
- Guy, H. I. & Evans, D. R. Cloning and expression of the mammalian multifunctional protein CAD in *Escherichia coli*. *J. Biol. Chem.* **269**, 23808–23816 (1994).
- Alessi, D. R., Cuenda, A., Cohen, P., Dudley, D. T. & Saltiel, A. R. PD 098059 is a specific inhibitor of the activation of mitogen-activated protein kinase kinase *in vitro* and *in vivo*. *J. Biol. Chem.* **270**, 27489–27494 (1995).
- Patterson, D. & Carnight, D. V. Biochemical genetic evidence analysis of pyrimidine biosynthesis in mammalian cells: I. Isolation of a mutant defective in the early steps of *de novo* pyrimidine synthesis. *Somat. Cell. Genet.* **3**, 483–495 (1977).
- Lazarowski, E. R., Homolya, L., Boucher, R. C. & Harden, T. K. Direct demonstration of mechanically induced release of cellular UTP and its implication for uridine nucleotide receptor activation. *J. Biol. Chem.* **272**, 24348–24354 (1997).
- Traut, T. W. Physiological concentrations of purines and pyrimidines. *Mol. Cell. Biochem.* **140**, 1–22 (1994).
- Carrey, E. A., Campbell, D. G. & Hardie, D. G. Phosphorylation and activation of hamster carbamyl phosphate synthetase II by cAMP-dependent protein kinase. A novel mechanism for regulation of pyrimidine nucleotide biosynthesis. *EMBO J.* **4**, 3735–3742 (1985).
- Shaw, S. M. & Carrey, E. A. Regulation of the mammalian carbamoyl-phosphate synthetase II by effectors and phosphorylation. Altered affinity for ATP and magnesium ions measured using the ammonia-dependent part reaction. *Eur. J. Biochem.* **207**, 957–965 (1992).
- Banerjee, L. C. & Davidson, J. N. Site-directed substitution of Ser 1406 of hamster CAD with glutamic acid alters the allosteric regulation of carbamyl phosphate synthetase II. *Som. Cell. Mol. Genet.* **23**, 37–49 (1997).
- Sahay, N., Guy, H. I., Liu, X. & Evans, D. R. Regulation of a *Escherichia coli*/mammalian chimeric carbamoyl-phosphate synthetase. *J. Biol. Chem.* **273**, 31195–31202 (1998).
- Li, X., Lee, J. W., Graves, L. M. & Earp, H. S. Angiotensin II stimulates ERK via two pathways in epithelial cells: protein kinase C suppresses a G-protein coupled receptor–EGF receptor trans-activation pathway. *EMBO J.* **17**, 2574–2583 (1998).
- Coleman, P. F., Suttle, D. P. & Stark, G. R. Purification from hamster cells of the multifunctional protein that initiates *de novo* synthesis of pyrimidine nucleotides. *J. Biol. Chem.* **252**, 6379–6385 (1977).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank J. Davidson for the gift of the G9c cells, and Y. He and R. Li for technical help. This work was supported by NIH grants to L.M.G. and D.R.E. and an ACS grant to H.I.G.

Correspondence and requests for materials should be addressed to L.M.G. (e-mail: lmg@med.unc.edu).

The joining of ribosomal subunits in eukaryotes requires eIF5B

Tatyana V. Pestova^{*,†}, Ivan B. Lomakin^{*}, Joon H. Lee[‡], Sang Ki Choi[‡], Thomas E. Dever[‡] & Christopher U. T. Hellen^{*}

^{*} Department of Microbiology and Immunology, State University of New York Health Science Center at Brooklyn, 450 Clarkson Avenue, Brooklyn, New York 11203, USA

[†] A.N. Belozersky Institute of Physico-Chemical Biology, Moscow State University, 119899 Moscow, Russia

[‡] Laboratory of Eukaryotic Gene Regulation, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892-2716, USA

Initiation of eukaryotic protein synthesis begins with the ribosome separated into its 40S and 60S subunits¹. The 40S subunit first binds eukaryotic initiation factor (eIF) 3 and an eIF2–GTP–initiator transfer RNA ternary complex. The resulting complex requires eIF1, eIF1A, eIF4A, eIF4B and eIF4F to bind to a messenger RNA and to scan to the initiation codon². eIF5 stimulates hydrolysis of eIF2-bound GTP and eIF2 is released from the 48S complex formed at the initiation codon before it is joined by a 60S subunit to form an active 80S ribosome^{3–8}. Here we show that hydrolysis of eIF2-bound GTP induced by eIF5 in 48S complexes is necessary but not sufficient for the subunits to join. A second factor termed eIF5B (relative molecular mass 175,000) is essential for this process. It is a homologue of the prokaryotic initiation factor IF2 (refs 6, 7) and, like it^{8–12}, mediates joining of subunits and has a ribosome-dependent GTPase activity that is essential for its function.

Recombinant eIF5 could not mediate joining of 60S subunits to 48S complexes assembled *in vitro* on globin mRNA, but a ribosomal salt wash (RSW) fraction contained the necessary additional activities (Fig. 1a). The active constituents were purified as apparently homogenous proteins of relative molecular mass 175,000 ($M_r = 175K$) (p175) and 58,000 (p58) (Fig. 1b) that were necessary and together, but not individually, sufficient for 80S complex formation on this mRNA (Fig. 1c). Recombinant eIF5 could replace p58 in this reaction (data not shown). The amino-terminal sequence of p175, GKKQKNKSEDSTKDDTDLGALAAEIEGAGAAK, identified it as a mouse homologue of prokaryotic IF2. Yeast (yIF2) and human (hIF2) homologues have also been identified^{6,7}. The translation defect in a *Saccharomyces cerevisiae* *fun12Δ* extract lacking yIF2 was rescued equally by p175 and recombinant glutathione S-transferase (GST)–ΔyIF2 and GST–ΔhIF2 fusion proteins lacking amino acids 1–395 and 1–586, respectively (Fig. 1d). GST–ΔhIF2 could replace native p175 in subunit joining (Fig. 1c). p175 is essential for subunits to join and we call it eIF5B because it acts with eIF5 in this process. Physical and biochemical properties suggest that eIF5B is a protein with $M_r \approx 150K$ that was previously implicated in subunit joining^{13–16} but subsequently discounted as an inactive contaminant of eIF5 (refs 3–5).

Previous analysis indicated that eIF5 and the ~150K protein had

overlapping and partially redundant functions in promoting hydrolysis of eIF2-bound GTP and subsequent joining of subunits^{3–5,13–16}, whereas we found that eIF5 and eIF5B were both necessary on native mRNA. These earlier analyses were mostly done using 48S complexes assembled on AUG triplets rather than on native mRNA. 48S complex formation on AUG triplets is much simpler than on native mRNA because it requires only 40S subunits, eIF2, GTP and initiator tRNA. To address this question we compared the activity of both proteins with different combinations of other factors using methionylpuromycin synthesis (which mimics formation of the first peptide bond) to assay formation of active 80S ribosomes on AUG triplets. Individually, eIF5 and eIF5B were equally active in a minimal system containing ribosomal subunits, eIF1A and eIF2, and their activities were not additive or synergistic (Fig. 2a). eIF1 enhanced the activity of eIF5 but not eIF5B, whereas eIF3 enhanced the activity of eIF5B and reduced the activity of eIF5. Together, eIF1 and eIF3 reduced the activity of eIF5B and, to a greater extent, that of eIF5. Methionylpuromycin synthesis was strongly

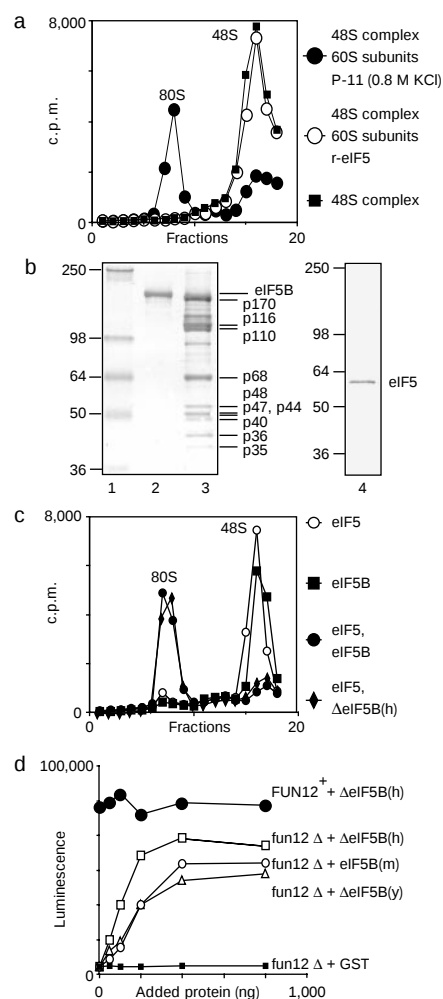


Figure 1 Formation of 80S complexes on β -globin mRNA requires eIF5 and eIF5B.

a, c, Sucrose density gradient centrifugation of 80S complexes assembled using recombinant eIF5 (r-eIF5), native eIF5B, recombinant human GST–eIF5B_{587–1220} (ΔeIF5B) and a RSW fraction (40–50% $(\text{NH}_4)_2\text{SO}_4$ cut of the 0.8 M KCl eluate from phosphocellulose) as indicated, 48S complexes and 60S subunits. 48S complexes were assembled using 40S subunits, GTP, Met-tRNA^{Met}, eIF1, eIF1A, eIF2, eIF3, eIF4A, eIF4B and eIF4F. Upper gradient fractions have been omitted. **b,** Purified native eIF5B (lane 2), subunits of eIF3 (lane 3) and eIF5 (lane 4) resolved by SDS-PAGE are indicated to the right. **d,** Translation activity in yeast *fun12Δ* extracts of luciferase mRNA incubated with the indicated amounts of native mouse eIF5B (m) or recombinant GST, yeast GST–eIF5B_{396–1002} (ΔeIF5B(y)) or human (ΔeIF5B(h)) proteins⁵.

increased by the synergistic activity of eIF5 and eIF5B in the presence of eIF1 and eIF3. These results are consistent with earlier reports but indicate that eIF5 and eIF5B are both required for joining of subunits in the presence of a complete set of factors.

eIF5B could not replace eIF2 in 48S complex formation on globin mRNA assayed by toe-printing analysis² (not shown) and could not replace eIF2 in methionylpuromycin synthesis (not shown) as was reported by yIF2 (ref. 6). Thus, despite its homology with IF2 and their common role in subunit joining, eIF5B differs from it in not promoting recruitment of Met-tRNA^{Met} to the small ribosomal subunit. This is instead done by eIF2, which has no prokaryotic homologue.

A requirement for both eIF5 and eIF5B for methionylpuromycin synthesis was apparent only in the presence of eIF1 and eIF3. Hydrolysis of GTP bound to 48S complexes is a prerequisite for subunit joining and was therefore compared with and without eIF1 and eIF3. GTP hydrolysis was stimulated equally by eIF5 and eIF5B when 48S complexes contained only eIF1A and eIF2. Both eIF1 and eIF3 reduced eIF5B's activity without affecting that of eIF5 (Fig. 2c, d). This effect could account for the reduced ability of eIF5B to promote methionylpuromycin synthesis in the presence of eIF1 and eIF3 (Fig. 2a). Hydrolysis of GTP induced by eIF5 in 48S complexes in the presence of a normal set of factors is therefore not sufficient for subunits to join, and eIF5B is required in these conditions. Hydrolysis of GTP in 48S complexes was induced considerably more weakly in these conditions by GST- Δ eIF5B than by native eIF5B (Fig. 2d). Native eIF5B and recombinant Δ eIF5B with a full set of factors had comparable subunit-joining activities on globin mRNA (Fig. 1c). We therefore compared the influence of different factors on the activity of these forms of eIF5B in the synthesis of methionylpuromycin. In the presence of eIF1A, eIF2 and ribosomal subunits but not eIF5, recombinant Δ eIF5B was 30% as active as intact native eIF5B in promoting synthesis of methionylpuromycin by 80S complexes assembled on AUG triplets (Fig. 2b). Their activities were almost identical when reactions also contained eIF1, eIF3 and eIF5. Normally, therefore, eIF5 triggers hydrolysis

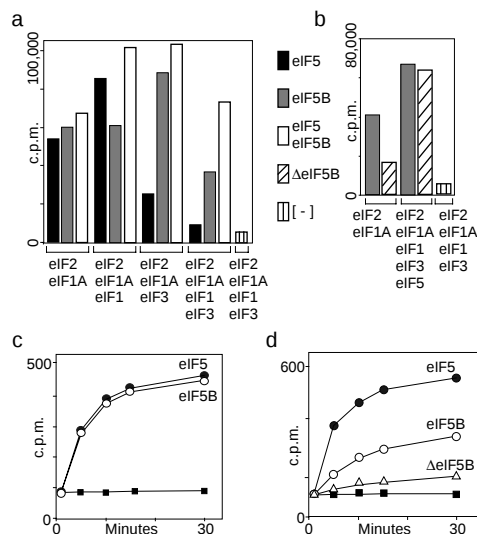


Figure 2 Dependence of the activities of eIF5 and eIF5B on the presence of eIF1 and eIF3. **a, b**, Methionylpuromycin synthesis by active 80S ribosomes assembled using recombinant eIF5, native eIF5B and GST-eIF5B₅₈₇₋₁₂₂₀ (Δ eIF5B), 48S complexes and 60S subunits. 48S complexes were assembled using 40S subunits, GTP, Met-tRNA^{Met}, AUG and factors as indicated. **c, d**, Activation by recombinant eIF5, native eIF5B or GST- Δ eIF5B of hydrolysis of eIF2-bound GTP in 48S complexes purified after assembly on AUG using 40S subunits, γ [³²P] GTP, Met-tRNA^{Met} and eIF1A and eIF2 (**c**) or eIF1, eIF1A, eIF2 and eIF3 (**d**). The data represent the average of three independent experiments with s.d. <10%.

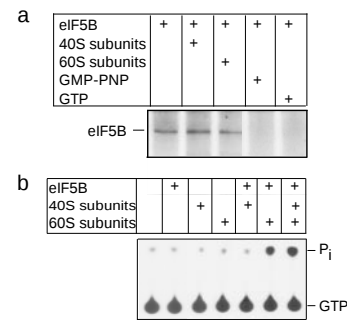


Figure 3 GTP binding and hydrolysis by eIF5B. **a**, Ultraviolet cross-linking of [³²P]GTP to native eIF5B in the presence of ribosomal subunits and unlabelled competitor nucleotides. Proteins resolved by SDS-PAGE were detected by autoradiography. **b**, Thin-layer chromatography analysis of stimulation by ribosomal subunits of GTP hydrolysis by eIF5B. The positions of [³²P]GTP, [³²P]Pi and eIF5B are indicated.

of eIF2-bound GTP. As full-length and N-terminally truncated eIF5B had different activities in inducing hydrolysis of eIF2-bound GTP, they probably have different affinities for components of the 48S complex.

eIF5B contains sequence motifs characteristic of GTP-binding proteins^{6,7}. eIF5B bound directly to [³²P]GTP in ultraviolet cross-linking assays independently of 40S or 60S subunits (Fig. 3a). Cross-linking of [³²P]GTP to eIF5B was outcompeted by unlabelled GTP, GDP or GMPPNP (a nonhydrolysable GTP analogue) (Fig. 3a), which shows that GTP bound to eIF5B exchanges readily. eIF5B had no detectable intrinsic GTPase activity; however, GTP hydrolysis was activated by 60S subunits, and much more by 60S and 40S subunits, but not by 40S subunits alone (Fig. 3b). GTP was hydrolysed by eIF5B in the presence of 40S and 60S subunits at a linear rate of ~1 pmol per min per μ g protein.

The ~150K subunit-joining factor has been reported to act catalytically^{14,16}. The activity of His₆- Δ eIF5B in promoting syn-

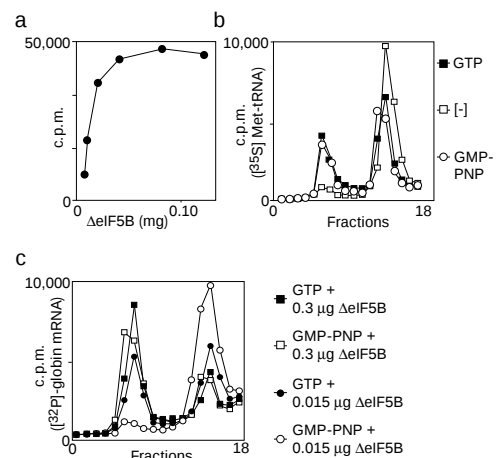


Figure 4 Catalytic activity of eIF5B. **a**, Methionylpuromycin synthesis in the presence of AUG, Met-tRNA^{Met}, 40S and 60S subunits, eIF1, eIF1A, eIF2, eIF5 and amounts of His₆- Δ eIF5B as indicated. **b, c**, Subunit joining in the presence of excess GMPPNP. **b**, Sucrose density gradient centrifugation of 80S complexes assembled using 60S subunits, eIF5, 1 mM GMPPNP or GTP if indicated, and 0.3 μ g His₆- Δ eIF5B added to gel-filtration purified 48S complexes pre-assembled using AUG, 40S subunits, 0.1 mM GTP, [³⁵S]Met-tRNA^{Met}, eIF1, eIF1A, eIF2 and eIF3. **c**, Sucrose density gradient centrifugation of 80S complexes assembled from 60S subunits, eIF5, 1 mM GMPPNP or GTP and amounts of His₆- Δ eIF5B as indicated added to reaction mixtures containing 48S complexes preassembled on [³²P]UTP-labelled globin mRNA from 40S subunits, 0.1 mM GTP, Met-tRNA^{Met}, eIF1, eIF1A, eIF2, eIF3, eIF4A, eIF4B and eIF4F.

thesis of methionylpuromycin was therefore assayed by titration in mixtures containing eIF1, eIF1A, eIF2, eIF3 and eIF5. Near-saturating amounts (0.35 pmol) of His₆-ΔeIF5B led to synthesis of ~1 pmol methionylpuromycin (Fig. 4a). eIF5B therefore acts in multiple cycles of subunit joining and eIF5B-GDP may not need a recycling factor. Binding and hydrolysis of GTP by His₆-ΔeIF5B could be required for it to adopt an active conformation or for the release of itself or other factors from assembled 80S ribosomes. His₆-ΔeIF5B promoted 80S ribosome assembly on β-globin mRNA (Fig. 4c) and on AUG triplets (not shown); however, it no longer acted catalytically when added with 1 mM GMPPNP-Mg²⁺, eIF5 and 60S subunits to 48S complexes that had been pre-assembled using 0.1 mM GTP. GTP was present with 10× excess GMPPNP so we cannot exclude that this activity is due to residual GTP-bound His₆-ΔeIF5B. To resolve this issue, we separated 48S complexes assembled on AUG in the presence of γ[³²P]-labelled GTP from unincorporated GTP by gel filtration. Similar amounts of 80S complex formed when pre-assembled 48S complexes were incubated with eIF5, His₆-ΔeIF5B, 60S subunits and GTP or GMPPNP (Fig. 4b). Hydrolysis of GTP is thus not essential for subunit joining, but stoichiometric rather than catalytic amounts of His₆-ΔeIF5B are required when GMPPNP is present. 80S complexes did not form when neither GTP nor GMPPNP were present. Thus, His₆-ΔeIF5B must bind GTP to adopt an active conformation but hydrolysis of bound GTP is not essential for joining of subunits.

We assembled 80S complexes on AUG in the presence of either GTP or GMPPNP from preformed (but not purified) 48S complexes (Fig. 5a), purified them by sucrose density gradient centrifugation and assayed synthesis of methionylpuromycin. About 60% of Met-tRNA^{Met} in 80S complexes assembled using GTP reacted with puromycin but only 8% of complexes assembled using GMPPNP were active (Fig. 5b). Synthesis of methionylpuromycin by purified 80S complexes pre-assembled using GTP was unaffected by the addition of GTP, GMPPNP or His₆-ΔeIF5B alone or His₆-ΔeIF5B and GTP, but was completely inhibited by His₆-ΔeIF5B and GMPPNP. These results may indicate that His₆-ΔeIF5B bound to GMPPNP can interact with pre-assembled 80S complexes, blocking their ability to react with puromycin. When 80S complexes assembled using GTP were resolved on sucrose gradients, no His₆-ΔeIF5B was detected in 40S, 48S, 60S or 80S complexes.

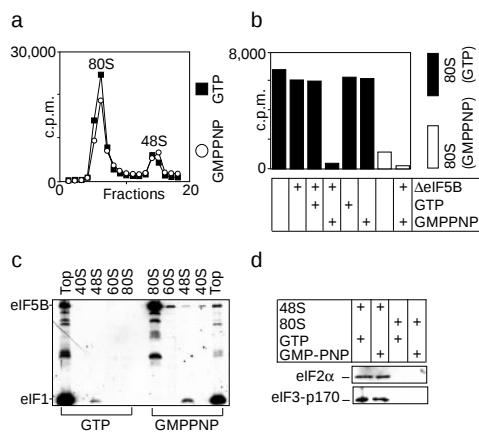


Figure 5 Hydrolysis of GTP bound to eIF5B is required for release of eIF5B from ribosomes. **a**, Sucrose density gradient centrifugation of 80S complexes assembled using 60S subunits, eIF5, 1 mM GTP or GMPPNP and 0.3 μg His₆-ΔeIF5B added to reaction mixtures containing 48S complexes assembled using AUG, 40S subunits, 0.1 mM GTP, [³⁵S]Met-tRNA^{Met}, eIF1, eIF1A, eIF2 and eIF3. **b**, Methionylpuromycin synthesis by 80S complexes purified from the gradients shown in **a** and incubated with 0.2 mM GTP, 0.2 mM GMPPNP and 0.3 μg His₆-ΔeIF5B as indicated. **c**, **d**, Immunodetection of ΔeIF5B, eIF1, eIF2α and eIF3 (p170) in ribosomal complexes isolated from these sucrose gradients.

However, a large amount of His₆-ΔeIF5B was bound to 80S complexes assembled using GMPPNP (Fig. 5c). The inability of His₆-ΔeIF5B to hydrolyse GMPPNP therefore locks the factor on 80S complexes and renders them inactive in methionylpuromycin synthesis. eIF1, eIF2 and eIF3 were detected in 48S but not 80S complexes assembled with either GTP or GMPPNP (Fig. 5c, d). Hydrolysis of GTP by His₆-ΔeIF5B is therefore not required to release these factors during subunit joining but is needed to release His₆-ΔeIF5B itself to form active ribosomes. The inability of eIF5B-GMPPNP to dissociate from 80S complexes explains the requirement for stoichiometric amounts of His₆-ΔeIF5B in assembly reactions in the presence of GMPPNP. Prokaryotic IF2 can also assemble ribosomes with GMPPNP and they are also inactive in methionylpuromycin synthesis because IF2 is not released^{11,12}. There are therefore significant similarities in the mechanism of action of eIF5B and its prokaryotic homologue IF2 in subunit joining. □

Methods

Purification of factors and ribosomal subunits

Ribosomes and factors were purified as described^{2,5,6}. Human liver eIF5 complementary DNA amplified by PCR was inserted into pET19b to express His₆-eIF5. eIF5 and eIF5B were purified from a Krebs-2 ascites RSW 40–50% (NH₄)₂SO₄ precipitation fraction, which was applied in buffer A (ref. 2) +100 mM KCl to a DEAE cellulose column and eluted with buffer A + 100 mM KCl and then +250 mM KCl. The latter fraction was dialysed against buffer A +100 mM KCl and applied to a phosphocellulose column. The active 400–80 mM KCl elution fraction was applied in buffer A +100 mM KCl to an FPLC MonoQ HR 5/5 column. Fractions were collected across a 100–500 mM KCl gradient. Apparently homogenous eIF5 eluted with 320 mM KCl and eIF5B with 400 mM KCl. The N-terminal sequence of eIF5B was obtained^{2,17} and used to obtain matching expressed sequence tags, which identified as eIF5B as an IF2 homologue^{6–8}.

Assembly and analysis of ribosomal complexes

We assembled 48S complexes incubating 0.5 μg [³²P]-labelled globin RNA (50,000–100,000 c.p.m. per μg RNA) for 5 min at 30 °C in 100 μl buffer B (20 mM Tris pH 7.5, 100 mM KAc, 2 mM DTT, 2.5 mM MgAc) with 1 mM ATP, 0.4 mM GTP, 0.25 mM spermidine, 5 pmol Met-tRNA, 5 pmol 40S subunits, 3 μg eIF2, 8 μg eIF3, 2 μg eIF4F, 2 μg eIF4A, 0.5 μg eIF4B, 0.5 μg eIF1 and 0.5 μg eIF1A. Pre-assembled 48S complexes were incubated with 5 pmol 60S subunits, 0.3 μg eIF5 and 1.5 μg eIF5B for 10 min to form 80S complexes. Complexes were resolved on sucrose density gradients⁵. To assay methionylpuromycin synthesis¹⁷, 40-μl mixtures of 2 pmol 40S subunits, 3 pmol [³⁵S]Met-tRNA (40,000 c.p.m. pmol⁻¹), 0.1 mM GTP, 1 nmol AUG triplet, 2 μg eIF2, 5 μg eIF3, 0.5 μg eIF1 and 0.5 μg eIF1A in different combinations were incubated at 37 °C for 5 min in buffer B. The Mg²⁺ concentration was elevated to 5 mM after adding 1 mM puromycin, 2.5 pmol 60S subunits, 0.3 μg eIF5 and 1 μg eIF5B. Incubation continued at 37 °C for 30 min.

GTP hydrolysis assays

We assembled 48S complexes by incubating 40 μl mixtures of buffer B, 50 μM GTP with 100 μCi γ[³²P]GTP, 1 nmol AUG, 2.5 pmol 40S subunits, 4 pmol Met-tRNA, 3 μg eIF2 and combinations of 5 μg eIF3, 0.5 μg eIF1 and 0.5 μg eIF1A for 5 min at 37 °C. Reaction mixtures were loaded on 2 ml Sepharose CL-6B columns pre-equilibrated with buffer B + 5 mM MgAc. Peak fractions of 48S complexes were incubated for 30 min at 37 °C after adding 0.3 μg eIF5 or 1 μg eIF5B or recombinant ΔeIF5B. 20-μl aliquots were removed and assayed for [³²P]Pi release using ammonium phosphomolybdate⁵. To analyse GTP hydrolysis by thin layer chromatography, we incubated 20-μl mixtures of buffer B, 1 μg native eIF5B, 3 pmol 40S subunits, 3 pmol 60S subunits and 50 μM GTP (with 20 μCi γ[³²P]GTP) at 37 °C for 20 min. Aliquots (2 μl) were spotted on polyethyleneimine-cellulose plates for chromatography done using 0.8 M LiCl/0.8 M acetic acid. [³²P]Pi release was determined⁵ using 5-μl aliquots.

Ultraviolet crosslinking of eIF5B to GTP

Mixtures (30 μl) of 2 μg eIF5B in buffer B and 100 μCi α-³²P]GTP with or without 3 pmol 40S and 60S subunits were incubated at 30 °C for 5 min, supplemented with unlabelled GTP or GMPPNP (to 1 mM) and incubated for 5 min more. Samples were irradiated on ice for 30 min.

Methionylpuromycin synthesis activity of purified 80S ribosomes

To assay methionylpuromycin synthesis by pre-assembled 80S ribosomes, 0.3 μg His₆-ΔeIF5B, eIF5, 60S subunits, 1 mM GTP-Mg²⁺ or GMPPNP-Mg²⁺ and 48S complexes pre-assembled on AUG with 0.1 mM GTP (but not purified) were incubated together and resolved on sucrose gradients. 48S and 80S peaks were determined by scintillation counting of 50-μl aliquots from 400-μl gradient fractions. 40S and 60S peaks were determined by optical density. Methionyl puromycin synthesis assays were done using 50-μl aliquots of 80S peak fractions diluted 10× in buffer B + 5 mM Mg²⁺ and incubated with 1 mM puromycin and combinations of 0.2 mM GTP, 0.2 mM GMPPNP and 0.3 μg His₆-ΔeIF5B.

Western blotting

Ribosomal complexes assembled and purified as described above were TCA-precipitated. Proteins were resolved on 12% polyacrylamide gel, transferred to nitrocellulose membrane and probed for eIF1 and eIF5B using T7-tag antibodies (Novagen) and for eIF2 α and eIF3 (p170) using specific antibodies.

Received 13 September; accepted 16 November 1999.

- Merrick, W. C. Mechanism and regulation of eukaryotic protein synthesis. *Microbiol. Rev.* **56**, 291–315 (1992).
- Pestova, T. V., Borukhov, S. I. & Hellen, C. U. T. Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature* **394**, 854–859 (1998).
- Chakrabarti, A. & Maitra, U. Functions of eukaryotic initiation factor 5 in the formation of an 80S ribosomal polypeptide chain initiation complex. *J. Biol. Chem.* **266**, 14039–14045 (1991).
- Das, K., Chesevich, J. & Maitra, U. Molecular cloning and expression of cDNA for mammalian translation initiation factor 5. *Proc. Natl Acad. Sci. USA* **90**, 3058–3062 (1993).
- Huang, H.-K., Yoon, H., Hannig, E. M. & Donahue, T. F. GTP hydrolysis controls stringent selection of the AUG start codon during translation initiation in *Saccharomyces cerevisiae*. *Genes Dev.* **11**, 2396–2413 (1997).
- Choi, S. K., Lee, J. H., Zoll, W. L., Merrick, W. C. & Dever, T. E. Promotion of Met-tRNA^{Met} binding to ribosomes by yIF2, a bacterial IF2 homolog in yeast. *Science* **280**, 1757–1760 (1998).
- Lee, J. H., Choi, S. K., Roll-Mecak, A., Burley, S. K. & Dever, T. E. Universal conservation in translation initiation revealed by human and archaeal homologs of bacterial translation factor IF2. *Proc. Natl Acad. Sci. USA* **96**, 4342–4347 (1999).
- Sacerdot, C., Dessen, P., Hershey, J. W. B., Plumbridge, J. A. & Grunberg-Manago, M. Sequence of the initiation factor IF2 gene; unusual protein features and homologies with elongation factors. *Proc. Natl Acad. Sci. USA* **81**, 7787–7791 (1984).
- Kolakofsky, D., Dewey, K. F., Hershey, J. W. B. & Thach, R. E. Guanosine 5'-triphosphatase activity of initiation factor f2. *Proc. Natl Acad. Sci. USA* **61**, 1066–1070 (1968).
- Godefroy-Colburn, T. et al. Light-scattering studies showing the effect of initiation factors on the reversible dissociation of *Escherichia coli* ribosomes. *J. Mol. Biol.* **94**, 461–478 (1975).
- Luchin, S. et al. *In vitro* study of two dominant inhibitory GTPase mutants of *Escherichia coli* translation initiation factor IF2. Direct evidence that GTP hydrolysis is necessary for factor recycling. *J. Biol. Chem.* **274**, 6074–6079 (1999).
- Lockwood, A. H., Sarkar, P. & Maitra, U. Release of polypeptide chain initiation factor IF-2 during initiation complex formation. *Proc. Natl Acad. Sci. USA* **69**, 3602–3605 (1972).
- Merrick, W. C., Kemper, W. M. & Anderson, W. F. Purification and characterization of homogenous initiation factor M2A from rabbit reticulocytes. *J. Biol. Chem.* **250**, 5556–5562 (1975).
- Trachsel, H., Emi, B., Schreier, M. H. & Staehelin, T. Initiation of mammalian protein synthesis. II. The assembly of the initiation complex with purified initiation factors. *J. Mol. Biol.* **116**, 755–767 (1977).
- Benne, R., Brown-Luedi, M. L. & Hershey, J. W. B. Purification and characterization of protein synthesis initiation factors eIF-1, eIF-4C, eIF-4D, and eIF-5 from rabbit reticulocytes. *J. Biol. Chem.* **253**, 3070–3077 (1978).
- Peterson, D. T., Safer, B. & Merrick, W. C. Role of eukaryotic initiation factor 5 in the formation of 80S initiation complexes. *J. Biol. Chem.* **254**, 7730–7735 (1979).
- Pestova, T. V., Shatsky, I. N., Fletcher, S. P., Jackson, R. J. & Hellen, C. U. T. A prokaryotic-like mode of binding of cytoplasmic eukaryotic ribosomes to the initiation codon during internal initiation of translation of Hepatitis C and Classical Swine fever virus RNAs. *Genes Dev.* **12**, 67–83 (1998).

Acknowledgements

We thank W. Merrick for discussions, D. Etchison and R. Schneider for antibodies, and L. Siconolfi-Baez for sequencing eIF5B. These studies were supported by grants from the NIH to C.U.T.H. and T.V.P.

Correspondence and requests for materials should be addressed to T.V.P. (e-mail: tpestova@netmail.hscbklyn.edu).

A synthetic oscillatory network of transcriptional regulators

Michael B. Elowitz & Stanislas Leibler

Departments of Molecular Biology and Physics, Princeton University, Princeton, New Jersey 08544, USA

Networks of interacting biomolecules carry out many essential functions in living cells¹, but the 'design principles' underlying the functioning of such intracellular networks remain poorly understood, despite intensive efforts including quantitative analysis of relatively simple systems². Here we present a complementary approach to this problem: the design and construction of a synthetic network to implement a particular function. We used three transcriptional repressor systems that are not part of any natural biological clock^{3–5} to build an oscillating network, termed

the repressilator, in *Escherichia coli*. The network periodically induces the synthesis of green fluorescent protein as a readout of its state in individual cells. The resulting oscillations, with typical periods of hours, are slower than the cell-division cycle, so the state of the oscillator has to be transmitted from generation to generation. This artificial clock displays noisy behaviour, possibly because of stochastic fluctuations of its components. Such 'rational network design' may lead both to the engineering of new cellular behaviours and to an improved understanding of naturally occurring networks.

In the network shown in Fig. 1a, the first repressor protein, LacI from *E. coli*, inhibits the transcription of the second repressor gene, *tetR* from the tetracycline-resistance transposon Tn10, whose protein product in turn inhibits the expression of a third gene, *cI* from λ phage. Finally, CI inhibits *lacI* expression, completing the cycle. That such a negative feedback loop can lead to temporal oscillations in the concentrations of each of its components can be seen from a simple model of transcriptional regulation, which we used to design the repressilator and study its possible behaviours (Box 1). In this model, the action of the network depends on several factors, including the dependence of transcription rate on repressor concentration, the translation rate, and the decay rates of the protein and messenger RNA. Depending on the values of these parameters, at least two types of solutions are possible: the system may converge toward a stable steady state, or the steady state may become unstable, leading to sustained limit-cycle oscillations (Fig. 1b, c).

We found that oscillations are favoured by strong promoters coupled to efficient ribosome-binding sites, tight transcriptional repression (low 'leakiness'), cooperative repression characteristics, and comparable protein and mRNA decay rates (Box 1, Fig. 1b). A general obstacle to the design of biochemical networks is uncertainty about the values of parameters that characterize the interactions between different components. In our network, estimates of the order of magnitude of the relevant parameters seem to be compatible with the possibility of oscillations. Nevertheless, to increase the chances that the artificial network would function in the oscillatory regime, we made two alterations to natural components. First, to address transcriptional strength and tightness, we used strong, yet tightly repressible hybrid promoters, developed previously, which combine the λ P_L promoter with *lac* and *tet* operator sequences⁶. Second, to bring the effective repressor protein lifetimes closer to that of mRNA (about 2 min, on average, in *E. coli*⁷), we inserted a carboxy-terminal tag, based on the *ssrA* RNA sequence⁸, at the 3' end of each repressor gene. Proteases in *E. coli* recognize this tag and target the attached protein for destruction^{9,10}. Such tags have been shown to reduce the half-life of the λ repressor DNA-binding domain from more than 60 min to around 4 min (ref. 8) and diminish the half-life of green fluorescent protein (GFP) to about 30–40 min (ref. 11).

With these considerations in mind, we used standard molecular biology techniques to construct a low-copy plasmid encoding the repressilator and a compatible, higher-copy reporter plasmid containing the tet-repressible promoter P_{tetO1} (ref. 6) fused to an intermediate stability variant of *gfp*¹¹ (Fig. 1a). Because the inducer IPTG interferes with repression by LacI, we expected that a transient pulse of IPTG might be capable of synchronizing a population of repressilator-containing cells. A culture of *E. coli* MC4100 containing the two plasmids and grown in media containing IPTG displayed what appeared to be a single damped oscillation of GFP fluorescence per cell after transfer to media lacking IPTG (results not shown). Because individual cells have no apparent means of maintaining synchronization, we studied the repressilator by isolating single cells under the microscope and monitoring their fluorescence intensity as they grew into small two-dimensional microcolonies consisting of hundreds of progeny cells. In these experiments, total observation time was limited by the colony entering a stationary phase after about 10 hours of growth at

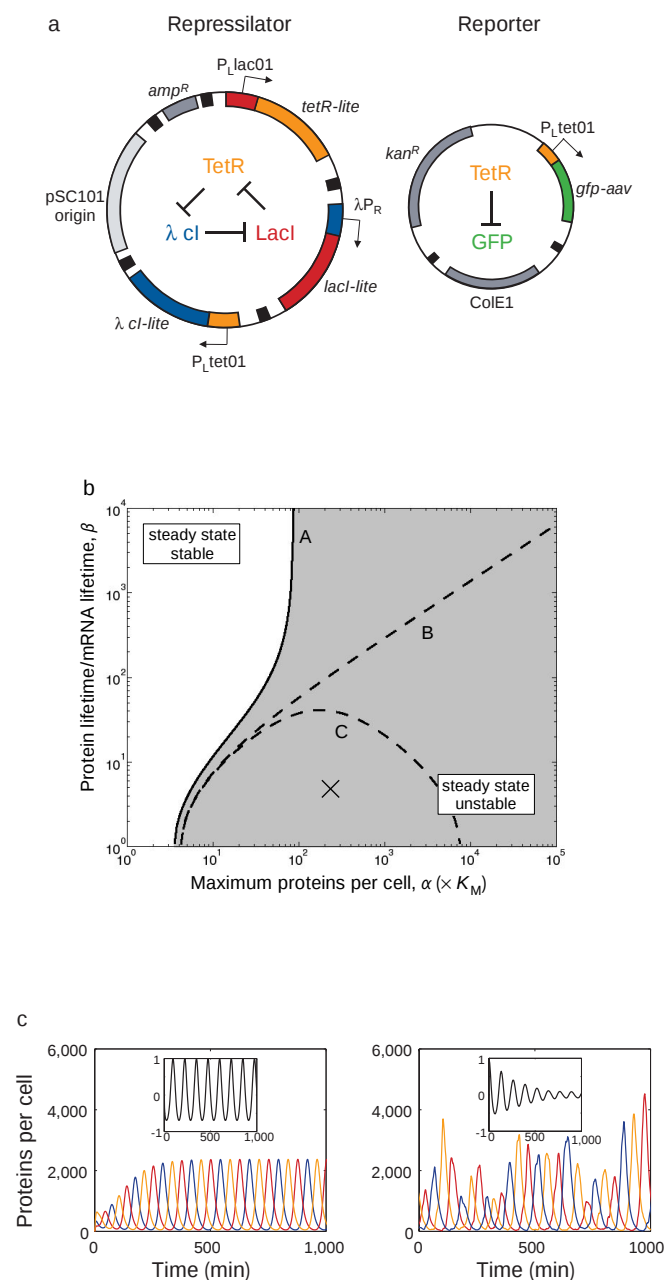


Figure 1 Construction, design and simulation of the repressilator. **a**, The repressilator network. The repressilator is a cyclic negative-feedback loop composed of three repressor genes and their corresponding promoters, as shown schematically in the centre of the left-hand plasmid. It uses P_{lacO1} and P_{tetO1} , which are strong, tightly repressible promoters containing *lac* and *tet* operators, respectively⁶, as well as P_R , the right promoter from phage λ (see Methods). The stability of the three repressors is reduced by the presence of destruction tags (denoted 'lite'). The compatible reporter plasmid (right) expresses an intermediate-stability GFP variant¹¹ (*gfp-aav*). In both plasmids, transcriptional units are isolated from neighbouring regions by T1 terminators from the *E. coli rrnB* operon (black boxes). **b**, Stability diagram for a continuous symmetric repressilator model (Box 1). The parameter space is divided into two regions in which the steady state is stable (top left) or unstable (bottom right). Curves A, B and C mark the boundaries between the two regions for different parameter values: A, $n = 2.1$, $\alpha_0 = 0$; B, $n = 2$, $\alpha_0 = 0$; C, $n = 2$, $\alpha_0/\alpha = 10^{-3}$. The unstable region (A), which includes unstable regions (B) and (C), is shaded. **c**, Oscillations in the levels of the three repressor proteins, as obtained by numerical integration. Left, a set of typical parameter values, marked by the 'X' in **b**, were used to solve the continuous model. Right, a similar set of parameters was used to solve a stochastic version of the model (Box 1). Colour coding is as in **a**. Insets show the normalized autocorrelation function of the first repressor species.

30 °C. At least 100 individual cell lineages in each of three micro-colonies were tracked manually, and their fluorescence intensity was quantified.

The timecourse of the fluorescence of one such cell is shown in Fig. 2. Temporal oscillations (in this case superimposed on an overall increase in fluorescence) occur with a period of around 150 minutes, roughly threefold longer than the typical cell-division time. The amplitude of oscillations is large compared with baseline levels of GFP fluorescence. At least 40% of cells were found to exhibit oscillatory behaviour in each of the three movies, as determined by a Fourier analysis criterion (see Methods). The range of periods, as estimated by the distribution of peak-to-peak intervals, is 160 ± 40 min (mean $\pm$ s.d., $n = 63$). After septation, GFP levels in the two sibling cells often remained correlated with one another for long periods of time (Fig. 3a–c). Based on the analysis of 179 septation events in the 3 movies, we measured an average half-time for sibling decorrelation of 95 ± 10 min, which is longer than the typical cell-division times of 50–70 min under these conditions. This indicates that the state of the network is transmitted to the progeny cells, despite a strong noise component. We observed significant variations in the period and amplitude of the oscillator output both from cell to cell (Fig. 3d), and over time in a single cell and its descendants (Fig. 3a–c). In some individuals, periods were omitted or phase delayed in one cell relative to its sibling (Fig. 3a, c).

Recent theoretical work has shown that stochastic effects may be responsible for noisy operation in natural gene-expression networks¹². Simulations of the repressilator that take into account the stochastic nature of reaction events and discreteness of network components also exhibit significant variability, reducing the correlation time for oscillations from infinity (in the continuous model) to about two periods (Box 1, Fig. 1c, insets). In general, we would like to distinguish such stochastic effects from possible intrinsically complex dynamics (such as intermittence or chaotic behaviour). Further studies are needed to identify and characterize the sources of fluctuations in the repressilator and other designed networks. In particular, longer experiments performed under chemostatic conditions should enable more complete statistical characterization of

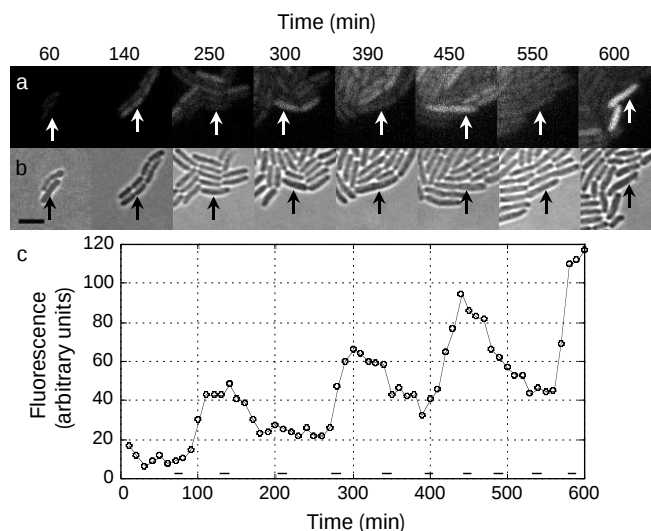


Figure 2 Repression in living bacteria. **a**, **b**, The growth and timecourse of GFP expression for a single cell of *E. coli* host strain MC4100 containing the repressilator plasmids (Fig. 1a). Snapshots of a growing microcolony were taken periodically both in fluorescence (**a**) and bright-field (**b**). **c**, The pictures in **a** and **b** correspond to peaks and troughs in the timecourse of GFP fluorescence density of the selected cell. Scale bar, 4 μ m. Bars at the bottom of **c** indicate the timing of septation events, as estimated from bright-field images.

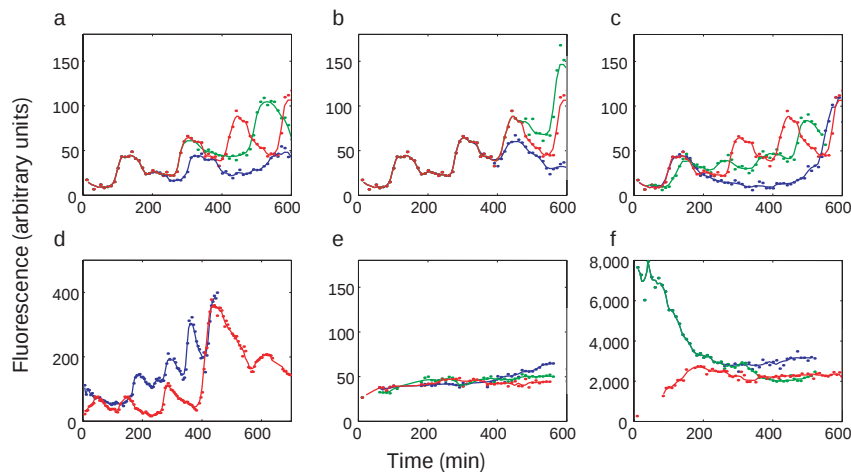


Figure 3 Examples of oscillatory behaviour and of negative controls. **a–c**, Comparison of the repressilator dynamics exhibited by sibling cells. In each case, the fluorescence timecourse of the cell depicted in Fig. 2 is redrawn in red as a reference, and two of its siblings are shown in blue and green. **a**, Siblings exhibiting post-septation phase delays relative to the reference cell. **b**, Examples where phase is approximately maintained but amplitude varies significantly after division. **c**, Examples of reduced period (green) and

long delay (blue). **d**, Two other examples of oscillatory cells from data obtained in different experiments, under conditions similar to those of **a–c**. There is a large variability in period and amplitude of oscillations. **e, f**, Examples of negative control experiments. **e**, Cells containing the repressilator were disrupted by growth in media containing 50 μ M IPTG. **f**, Cells containing only the reporter plasmid.

Box 1

Network design

Design of the repressilator started with a simple mathematical model of transcriptional regulation. We did not set out to describe precisely the behaviour of the system, as not enough is known about the molecular interactions inside the cell to make such a description realistic. Instead, we hoped to identify possible classes of dynamic behaviour and determine which experimental parameters should be adjusted to obtain sustained oscillations.

Deterministic, continuous approximation

Three repressor-protein concentrations, p_i , and their corresponding mRNA concentrations, m_i (where i is *lacI*, *tetR* or *cl*) were treated as continuous dynamical variables. Each of these six molecular species participates in transcription, translation and degradation reactions. Here we consider only the symmetrical case in which all three repressors are identical except for their DNA-binding specificities. The kinetics of the system are determined by six coupled first-order differential equations:

$$\begin{aligned} \frac{dm_i}{dt} &= -m_i + \frac{\alpha}{(1+p_i^n)} + \alpha_0 & (i = lacI, tetR, cl) \\ \frac{dp_i}{dt} &= -\beta(p_i - m_i) & (j = cl, lacI, tetR) \end{aligned}$$

where the number of protein copies per cell produced from a given promoter type during continuous growth is α_0 in the presence of saturating amounts of repressor (owing to the 'leakiness' of the promoter), and $\alpha + \alpha_0$ in its absence; β denotes the ratio of the protein decay rate to the mRNA decay rate; and n is a Hill coefficient. Time is rescaled in units of the mRNA lifetime; protein concentrations are written in units of K_M , the number of repressors necessary to half-maximally repress a promoter; and mRNA concentrations are rescaled by their translation efficiency, the average number of proteins produced per mRNA molecule. The numerical solution of the model shown in Fig. 1c used the following parameter values: promoter strength, 5×10^{-4} (repressed) to 0.5 (fully induced) transcripts per s; average translation efficiency, 20 proteins per transcript, Hill coefficient, $n = 2$; protein half-life, 10 min; mRNA half-life, 2 min; K_M , 40 monomers per cell.

This system of equations has a unique steady state, which becomes unstable when $\frac{(\beta+1)^2}{\beta} < \frac{3X^2}{4+2X}$, where $X = -\frac{\alpha n p^{n-1}}{(1+p^n)^2}$ and p is the solution to $p = \frac{\alpha}{1+p^n} + \alpha_0$. The boundary between the stable and

unstable domains can therefore be plotted (Fig. 1b). The unstable domain becomes much larger when the Hill coefficient increases, removing any limitation on β for sufficiently large α (compare curve B, for which $n = 2$, to curve A, for which $n = 2.1$). The effect of leakiness, α_0 , can be seen by plotting the stability boundary for a constant ratio of α_0/α , as in curve C. When α_0 becomes comparable to K_M (1 in our units), the unstable domain shrinks (compare curve B, for which $\alpha_0 = 0$, to curve C, for which $\alpha_0/\alpha = 10^{-3}$). Similar analysis of the stability of the steady state can be performed for generalized models of cyclic transcriptional feedback loops. The simplest such networks supporting limit-cycle oscillations are those containing a single repressor and a single activator, or an odd number of repressors exceeding 3. In general, the period of oscillations in such networks is determined mainly by the protein stability¹⁶. More detailed calculations, with non-Hill-function repression curves, or using thermodynamic binding energies to predict equilibrium operator occupancies, and taking repressor dimerization into account, yield similar stability results¹⁹. It is possible that, in addition to simple oscillations, this and more realistic models may exhibit other complex types of dynamic behaviour.

Stochastic, discrete approximation

The preceding analysis neglects the discrete nature of the molecular components and the stochastic character of their interactions, however. Such effects are believed to be important in biochemical and genetic networks¹². We therefore adapted the above equations to perform stochastic simulations, as described²⁰. To obtain cooperativity in repression analogous to the continuous case, we assumed the presence of two operator sites on each promoter and the following reactions: binding of proteins to each operator site ($1 \text{ nM}^{-1} \text{ s}^{-1}$); unbinding of protein from the first-occupied (224 s^{-1}) and the second-occupied (9 s^{-1}) operator; transcription from occupied ($5 \times 10^{-4} \text{ s}^{-1}$) and from unoccupied (0.5 s^{-1}) promoters; translation ($0.167 \text{ mRNA}^{-1} \text{ s}^{-1}$); protein decay (10 min half-life); and mRNA decay (2 min half-life). These parameters were chosen to correspond as closely as possible to the continuous model described above, assuming that 1 molecule per cell corresponds to a concentration of $\sim 1 \text{ nM}$. Oscillations persist for these parameter values (Fig. 1c) but with a large variability, resulting in a finite autocorrelation time (compare insets in Fig. 1c).

the noisy repressilator dynamics. In addition, varying the host species and genetic background would allow us to check for, and minimize, spurious interactions with endogenous cellular sub-systems, and to investigate how the network is embedded in the cell. For instance, in the repressilator network, the cell-division cycle does not seem to be coupled with the repressilator, as the timing of oscillations is uncorrelated with cell septation events (Fig. 2). However, entry into the stationary phase causes the repressilator to halt, indicating that the network is coupled to the global regulation of cell growth.

The levels of many cellular components vary over time in growing cells, and even strains that constitutively express GFP exhibit significant heterogeneity, so we performed several control experiments to check that the observed oscillations are indeed due to the repressilator (Fig. 3e, f). These included deliberate disruption of the network (by adding sufficient IPTG to interfere with LacI) and observation of GFP expression in the absence of the repressilator (from plasmids with different promoters and origins of replication).

Our results show that it is possible to design and construct an artificial genetic network with new functional properties from generic components that naturally occur in other contexts. Such work is analogous to the rational design of functional proteins from well-characterized motifs¹³. Further characterization of components and alteration of network connectivity may reveal general features of this and related networks, and provide a basis for improved design and possible use in biotechnological applications. Moreover, comparing designed networks with their evolved counterparts may also help us to understand the 'design principles' underlying the latter. For instance, circadian clocks are found in many organisms, including cyanobacteria in which the cell-division time may be shorter than the period^{14–15}. However, the reliable performance of such circadian oscillators can be contrasted with the noisy, variable behaviour of the repressilator. Instead of three repressors, it seems that circadian oscillators use both positive and negative control elements. Does this design lead to improved reliability? Recent theoretical analysis suggests that, in the presence of interactions between positive and negative control elements that lead to bistable, hysteretic behaviour, an oscillating circuit does indeed exhibit high noise-resistance¹⁶. It would be interesting to see whether one could build an artificial analogue of the circadian clock, and, if so, whether such an analogue would display the noise resistance and temperature compensation of its natural counterpart. □

Methods

Network construction

The three repressors (*lacI*, *tetR* and *cl*), the λ promoter P_R and the unstable variants of *gfp*¹¹ were all cloned by the polymerase chain reaction (PCR) and verified by sequencing. In the case of *lacI*, the GTG start codon was changed to ATG with the 5' primer. 3' PCR primers were used to add the coding sequence for the 11-amino-acid *ssrA* tag to the three repressors⁸. Tagging of full-length λ repressor, as well as *lacI* and *tetR*, resulted in functional proteins that required increased induction to achieve full repression, consistent with decreased stability. Promoter sequences P_{lacO1} and P_{tetO1} were obtained from pZE21-MCS1 and pZE12-luc⁶. The repressilator plasmid was constructed in three successive cloning steps from intermediate plasmids containing the three transcriptional units. The reporter plasmid was constructed by cloning the GFPaav variant¹¹, which has a half-life of around 90 min, into pZE21-MCS1 (ref. 6). The finite rate at which GFP is oxidized to its fluorescent form¹⁷ and its relatively long half-life limit our ability to observe fast dynamic phenomena. The presence of additional *tet* operators on the reporter plasmid constitutes a perturbation to the single-plasmid repressilator network by titrating out repressors¹⁸.

Data acquisition and analysis

Cells of *E. coli lac*⁻ strain MC4100 (unless otherwise specified) transformed with appropriate plasmids were grown in minimal media (7.6 mM [NH₄]₂SO₄, 2 mM MgSO₄, 30 μ M FeSO₄, 1 mM EDTA, 60 mM potassium phosphate, pH 6.8) supplemented with 0.5% glycerol, 0.1% casamino acids and appropriate antibiotics (20 μ g ml⁻¹ kanamycin or 20 μ g ml⁻¹ ampicillin), when needed. Time-lapse microscopy was conducted on a Zeiss Axiovert 135TV microscope equipped with a 512 $\times$ 512-pixel cooled CCD camera (Princeton Instruments). Cultures were grown for at least 10 hours to optical density of 0.1 at 600 nm, diluted into fresh media, spotted between a coverslip and 1 ml of liquid 2%

SeaPlaque low-melt agarose (FMC) in media, and sealed. The temperature of the samples was maintained at 30–32 °C by using Peltier devices (Melcor). Bright-field (0.1 s) and epifluorescence (0.05–0.5 s) exposures were taken periodically (every 5 or 10 min). All light sources (standard 100 W Hg and halogen lamps) were shuttered between exposures. Images were flat-field corrected with custom software. For the synchronization experiment, overnight cultures were diluted back 1:100 in media containing 100 μ M IPTG, grown to mid-log, washed several times, and diluted again and grown in 96-well plates at 30 with shaking in a Wallac Victor 2 plate reader, while fluorescence and absorbance measurements were taken every 5 min.

For analysis, cells were selected from the final frames of bright-field movies, without regard to their fluorescence signals, and tracked manually backward in time until the first frame. At each time point, the cell position was identified on the bright-field image, and fluorescence intensity data were averaged over a 28-pixel region, similar in size to the cell diameter, in the corresponding location on the fluorescence image. A fast Fourier transform was applied to the temporal fluorescence signal from each analysed cell lineage and divided by the transform of a decaying exponential with a time constant of 90 min, the measured lifetime of GFPaav. Power spectra exhibiting peaks of more than four times the background at frequencies of 0.2–0.5 per hour were classified as oscillatory (the choice of threshold alters the fraction of oscillatory cells defined by this criterion). The sibling 'decorrelation' half-time was defined as the time necessary for the quantity $|I_1(t) - I_2(t)|/(I_1(t) + I_2(t))$, averaged over pairs of daughter cells, to reach half of its asymptotic value. Here, $I_1(t)$ and $I_2(t)$ denote the fluorescence intensities of the two sibling cells at times, t , starting from the moment of septation (± 10 min). In this analysis, only cells judged to be oscillatory by the Fourier criterion were considered.

Various negative control experiments were performed. First, 50 μ M IPTG was added to the media to disrupt the functioning of the network (Fig. 3e). Second, we used a version of the repressilator plasmid lacking all but the *tetR* transcriptional unit (in *lacI*^R strain JM109). We thus varied GFP expression of the reporter plasmid by controlling TetR levels with 0–20 μ M IPTG. Third, we examined cells containing only the reporter plasmid (Fig. 3f). Fourth, we measured GFP expression from reporter plasmids modified in several ways either by replacing the P_{tetO1} promoter with each of the two other promoters, and *gfp-aav* with *gfp-lite*¹¹ (the suffix 'lite' indicates the presence of a C-terminal *ssrA* tag), or by replacing the ColE1 origin of replication with the lower-copy pSC101 origin normally used on the repressilator plasmid. In none of these control experiments did we observe oscillations similar to those produced by the repressilator.

Received 6 July; accepted 9 November 1999.

- Bray, D. Protein molecules as computational elements in living cells. *Nature* **376**, 307–312 (1995).
- Koshland, D. E. Jr The era of pathway quantification. *Science* **280**, 852–853 (1998).
- Winfree, A. T. *The Geometry of Biological Time* (Springer, Berlin, 1990).
- Goldbeter, A. *Biochemical Oscillations and Cellular Rhythms* (Cambridge Univ. Press, 1996).
- Thomas, R. & D'Ari, R. *Biological Feedback* (CRC Press, Boca Raton, 1990).
- Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/11-12 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
- Kushner, S. R. in *Escherichia Coli and Salmonella: Cellular and Molecular Biology* (ed. Neidhardt, F. C.) (ASM, Washington DC, 1996).
- Keiler, K. C., Waller, P. R. & Sauer, R. T. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271**, 990–993 (1996).
- Gottesman, S., Roche, E., Zhou, Y. & Sauer, R. T. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev.* **12**, 1338–1347 (1998).
- Herman, C., Thevenet, D., Boulloc, P., Walker, G. C. & D'Ari, R. Degradation of carboxy-terminal-tagged cytoplasmic proteins by the *Escherichia coli* protease HflB (FtsH). *Genes Dev.* **12**, 1348–1355 (1998).
- Andersen, J. B. et al. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol.* **64**, 2240–2246 (1998).
- McAdams, H. H. & Arkin, A. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* **15**, 65–69 (1999).
- Bryson, J. W. et al. Protein design: a hierarchical approach. *Science* **270**, 935–941 (1995).
- Dunlap, J. C. Molecular bases for circadian clocks. *Cell* **96**, 271–290 (1999).
- Kondo, T. et al. Circadian rhythms in rapidly dividing cyanobacteria. *Science* **275**, 224–227 (1997).
- Barkai, N. & Leibler, S. Circadian clocks limited by noise. *Nature* (submitted????). (Eds to update)
- Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544 (1998).
- Glascok, C. B. & Weickert, M. J. Using chromosomal lacI^{Q1} to control expression of genes on high-copy-number plasmids in *Escherichia coli*. *Gene* **223**, 221–231 (1998).
- Elowitz, M. B. *Transport, Assembly, and Dynamics in Systems of Interacting Proteins*. Thesis, Princeton Univ., Princeton (1999).
- Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).

Acknowledgements

We thank H. Bujard, S. Freundlieb, A. Hochschild, R. Lutz and C. Sternberg for plasmids and advice; U. Alon, N. Barkai, P. Cluzel, L. Frisen, C. Guet, T. Hyman, R. Kishony, A. Jaedicke, P. Lopez, F. Nédélec, S. Pichler, R. Kishony, T. Silhavy, T. Surrey, J. Vilar, C. Wiggins and E. Winfree for discussions; M. Surette for advice and encouragement; L. Hartwell and C. Weitz for comments on the manuscript; and F. Kafatos and E. Karsenti for hospitality and support at the European Molecular Biology Laboratory (EMBL), where part of this work was done. This work was partly supported by the US National Institutes of Health and the von Humboldt Foundation.

Correspondence and requests for materials should be addressed to M.B.E. (e-mail: melowitz@princeton.edu).

Construction of a genetic toggle switch in *Escherichia coli*

Timothy S. Gardner^{*†}, Charles R. Cantor^{*} & James J. Collins^{*†}

^{*} Department of Biomedical Engineering, [†] Center for BioDynamics and [‡] Center for Advanced Biotechnology, Boston University, 44 Cummings Street, Boston, Massachusetts 02215, USA

It has been proposed¹ that gene-regulatory circuits with virtually any desired property can be constructed from networks of simple regulatory elements. These properties, which include multistability and oscillations, have been found in specialized gene circuits such as the bacteriophage λ switch² and the *Cyanobacteria* circadian oscillator³. However, these behaviours have not been demonstrated in networks of non-specialized regulatory components. Here we present the construction of a genetic toggle switch—a synthetic, bistable gene-regulatory network—in *Escherichia coli* and provide a simple theory that predicts the conditions necessary for bistability. The toggle is constructed from any two repressible promoters arranged in a mutually inhibitory network. It is flipped between stable states using transient chemical or thermal induction and exhibits a nearly ideal switching threshold. As a practical device, the toggle switch forms a synthetic, addressable cellular memory unit and has implications for biotechnology, biocomputing and gene therapy.

The design and construction of synthetic gene-regulatory networks would be greatly facilitated by a theory with predictive capability. Previous work using a reconstituted enzyme system⁴ showed that nonlinear mathematics can predict qualitative behaviours of biochemical reaction networks, including multistability and hysteresis. However, a practical theory of gene-regulatory networks has lagged behind that of enzyme regulatory networks. A variety of physical and mathematical approaches, including logical (discrete)^{5–10}, piecewise linear¹¹, nonlinear^{12–14}, statistical–mechanical^{15,16} and stochastic^{17–19} formulations of the underlying biochemical dynamics, have been used in the past. Owing to the difficulty of testing their predictions, these theories have not, in general, been verified experimentally. Here we have integrated theory and experiment by constructing and testing a synthetic, bistable gene circuit based on the predictions of a simple mathematical model.

The toggle switch is composed of two repressors and two constitutive promoters (Fig. 1). Each promoter is inhibited by the repressor that is transcribed by the opposing promoter. We selected this design for the toggle switch because it requires the fewest genes and *cis*-regulatory elements to achieve robust bistable behaviour. By robust, we mean that the toggle exhibits bistability over a wide range of parameter values and that the two states are tolerant of the fluctuations inherent in gene expression (the toggle switch will not flip randomly between states). Although bistability is theoretically possible with a single, autocatalytic promoter, it would be less

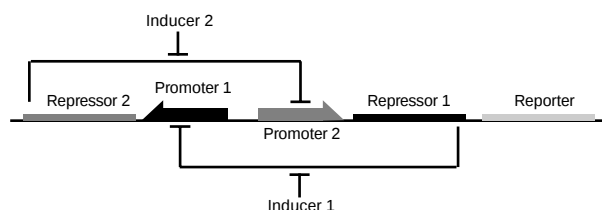


Figure 1 Toggle switch design. Repressor 1 inhibits transcription from Promoter 1 and is induced by Inducer 1. Repressor 2 inhibits transcription from Promoter 2 and is induced by Inducer 2.

robust and more difficult to tune experimentally. In addition, the chosen toggle design does not require any specialized promoters, such as the P_R/P_{RM} promoter of bacteriophage λ . Bistability is possible with any set of promoters and repressors as long as they fulfil the minimum set of conditions described in Box 1 and Fig. 2.

The bistability of the toggle arises from the mutually inhibitory arrangement of the repressor genes. In the absence of inducers, two stable states are possible: one in which promoter 1 transcribes repressor 2, and one in which promoter 2 transcribes repressor 1. Switching is accomplished by transiently introducing an inducer of the currently active repressor. The inducer permits the opposing repressor to be maximally transcribed until it stably represses the originally active promoter.

All toggle switches are implemented on *E. coli* plasmids conferring ampicillin resistance and containing the pBR322 ColE1 replication origin. The toggle switch genes are arranged as a type IV plasmid, as shown in Fig. 3. Although all genes and promoters are

Box 1

The toggle model

The behaviour of the toggle switch and the conditions for bistability can be understood using the following dimensionless model for the network:

$$\frac{du}{dt} = \frac{\alpha_1}{1 + v^\beta} - u \quad (1a)$$

$$\frac{dv}{dt} = \frac{\alpha_2}{1 + u^\gamma} - v \quad (1b)$$

where u is the concentration of repressor 1, v is the concentration of repressor 2, α_1 is the effective rate of synthesis of repressor 1, α_2 is the effective rate of synthesis of repressor 2, β is the cooperativity of repression of promoter 2 and γ is the cooperativity of repression of promoter 1. The above model is derived from a biochemical rate equation formulation of gene expression^{24–27}. The final form of the toggle equations preserves the two most fundamental aspects of the network: cooperative repression of constitutively transcribed promoters (the first term in each equation), and degradation/dilution of the repressors (the second term in each equation).

The parameters α_1 and α_2 are lumped parameters that describe the net effect of RNA polymerase binding, open-complex formation, transcript elongation, transcript termination, repressor binding, ribosome binding and polypeptide elongation. The cooperativity described by β and γ can arise from the multimerization of the repressor proteins and the cooperative binding of repressor multimers to multiple operator sites in the promoter. An additional modification to equation (1) is needed to describe induction of the repressors (Fig. 5).

The geometric structure of equation (1), illustrated in Fig. 2a and b, reveals the origin of the bistability: the nullclines ($du/dt = 0$ and $dv/dt = 0$ in Fig. 2) intersect at three points, producing one unstable and two stable steady states. From Fig. 2a and b, three key features of the system become apparent. First, the nullclines intersect three times because of their sigmoidal shape, which arises for $\beta, \gamma > 1$. Thus, the bistability of the system depends on the cooperative repression of transcription. Second, the rates of synthesis of the two repressors must be balanced. If the rates are not balanced, the nullclines will intersect only once, producing a single stable steady state. This situation arises in plasmid pKE105. Third, the structure of the toggle network creates two basins of attraction. Thus, a toggle with an initial condition anywhere above the separatrix will ultimately settle to state 1, whereas a toggle starting below the separatrix will settle to state 2.

The conditions for a bistable toggle network are illustrated in Fig. 2c and d. As the rates of repressor synthesis are increased, the size of the bistable region increases. Furthermore, the slopes of the bifurcation lines, for large α_1 and α_2 , are determined by β and γ . Thus, to obtain bistability, at least one of the inhibitors must repress expression with cooperativity greater than one. Moreover, higher order cooperativity will increase the robustness of the system, allowing weaker promoters to achieve bistability and producing a broader bistable region.

To investigate the conditions required for bistability, six variants of the toggle switch (four pTAK plasmids and two pIKE plasmids) were constructed by inserting RBS sequences of differing strengths into the RBS1 site. All four pTAK plasmids exhibited bistability, whereas only one pIKE plasmid (pIKE107) exhibited bistability. The demonstration of bistability is illustrated in Fig. 4. In this experiment, the toggle and control plasmids were grown in *E. coli* strain

The ideal threshold, or bifurcation, in the pTAK117 toggle switch is illustrated both theoretically and experimentally in Fig. 5. In this experiment, pTAK117 (initially in the low state) and pTAK102 (as a control) were grown in 13 different concentrations of IPTG for 17 h

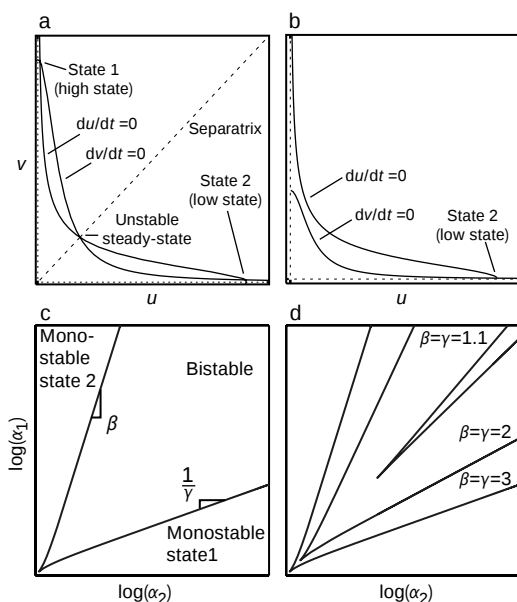


Figure 2 Geometric structure of the toggle equations. **a**, A bistable toggle network with balanced promoter strengths. **b**, A monostable toggle network with imbalanced promoter strengths. **c**, The bistable region. The lines mark the transition (bifurcation) between bistability and monostability. The slopes of the bifurcation lines are determined by the exponents β and γ for large α_1 and α_2 . **d**, Reducing the cooperativity of repression (β and γ) reduces the size of the bistable region. Bifurcation lines are illustrated for three different values of β and γ . The bistable region lies inside of each pair of curves.

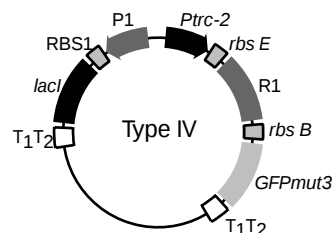


Figure 3 The toggle switch plasmid. Promoters are marked by solid rectangles with arrowheads. Genes are denoted with solid rectangles. Ribosome binding sites and terminators (T_1 T_2) are denoted by outlined boxes. Different P1 promoters, RBS1 ribosome binding sites, and/or R1 repressors, are used for the various toggle switches. Plasmid types I–III, used in the construction and testing of the toggle components, are described in the Supplementary Information.

to steady state, being diluted twice (at 6 and 12.5 h) into fresh medium with the same IPTG concentration. Induction of the pTAK102 control has the familiar sigmoidal shape. In contrast, the pTAK117 toggle follows the induction curve of pTAK102 up to an IPTG concentration of 40 μ M, at which point it crosses the bifurcation and exhibits a quasi-discontinuous jump to the high expression state. Owing to the natural fluctuations in gene expression, the bifurcation is not a perfect discontinuity as predicted by the deterministic toggle equations. The stochastic nature of gene expression causes variability in the location of the switching threshold and thus blurs the bifurcation point. Near the bifurcation, this blurriness is realized as a bimodal distribution of cells (Fig. 5c).

The switching time of the pTAK117 plasmid from the low to the high state and from the high to the low state is shown in Fig. 6. In this experiment, pTAK117 cells in the low state were diluted in fresh medium and induced with 2 mM IPTG. Separate cultures were grown for 35 min to 6 h before being washed and diluted in fresh medium with no inducer. Growth was continued until 10.25 h after the start of the experiment and cells were assayed in the flow cytometer. Conversely, pTAK117 cells in the high state were diluted in fresh medium with no inducer. Separate cultures were grown at $41 \pm 1^\circ\text{C}$ for 35 min to 6 h before being diluted in fresh medium with no inducer. Growth was continued at standard temperature until 10.25 h after the start of the experiment and cells were assayed in the flow cytometer.

As shown by the appearance of a bimodal distribution at 4 h (Fig. 6), the pTAK117 plasmid begins switching to the high state after 3–4 h of IPTG induction. By 5 h the switching is nearly complete, and by 6 h it is complete. On the other hand, switching from the high state to the low state is completed in 35 min or less. The primary determinant of switching time is the rate of elimination of the repressor proteins. Switching from low to high requires

the gradual dilution, by cell growth, of the IPTG-bound Lac repressor. On the other hand, switching from high to low is effected by immediate thermal destabilization of the temperature-sensitive λ repressor. Thus, switching to the low state is substantially faster than switching to the high state. Furthermore, the configuration of the pTAK117 plasmid—the rate of Lac repressor synthesis is more than an order of magnitude higher than the rate of λ repressor synthesis—suggests that the low state is more stable than the high state.

Our approach to the construction of a genetic toggle switch represents a significant departure from traditional genetic engineering in that we rely primarily on the manipulation of network

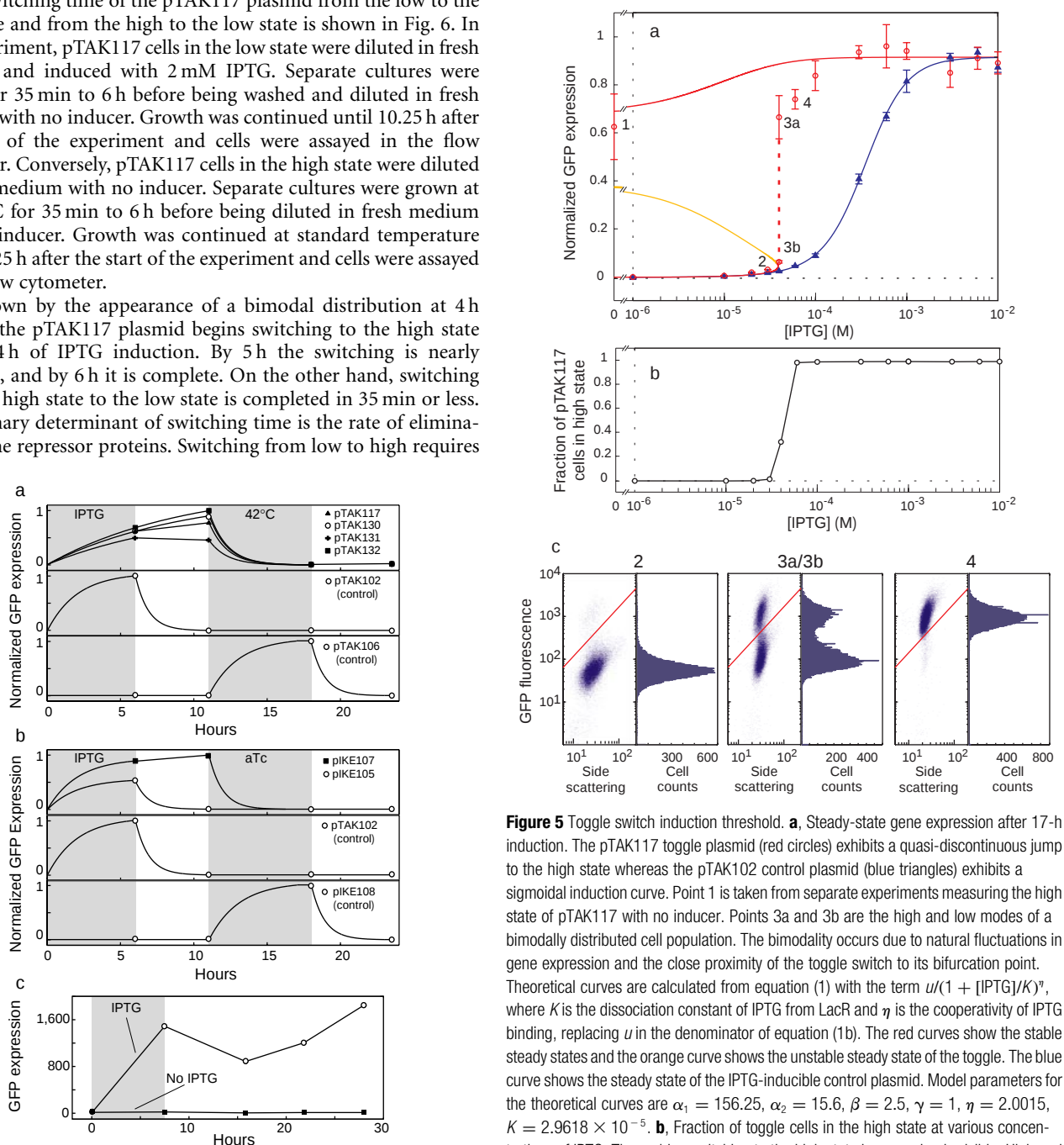


Figure 4 Demonstration of bistability. The grey shading indicates periods of chemical or thermal induction. The lines in **a** and **b**, which are approximations of the switching dynamics, are included for clarity. **a**, pTAK toggle plasmids and controls. **b**, pLac toggle plasmids and controls. **c**, Demonstration of the long-term stability of the separate expression states in pTAK117.

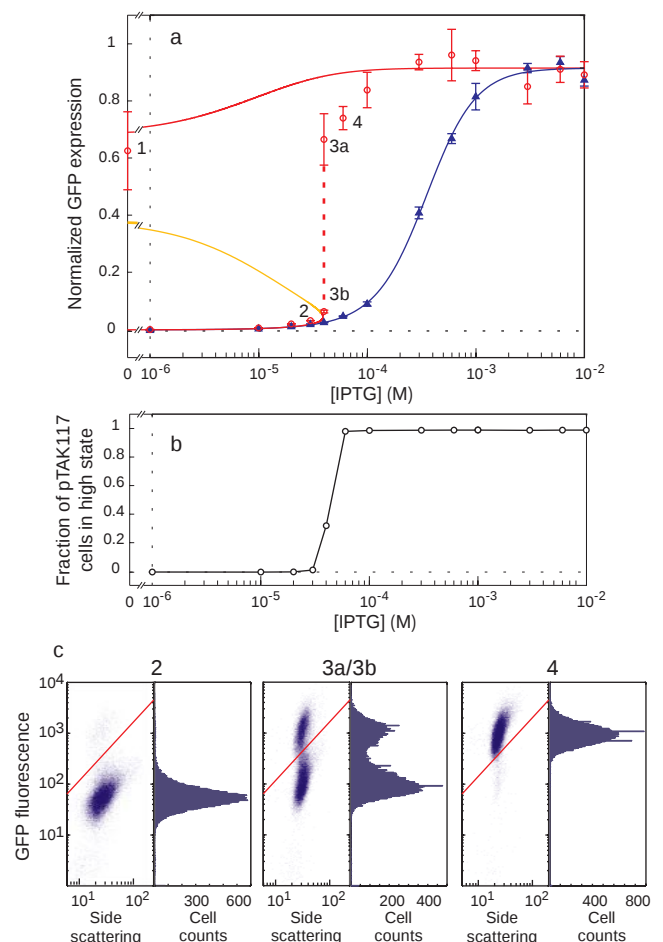


Figure 5 Toggle switch induction threshold. **a**, Steady-state gene expression after 17-h induction. The pTAK117 toggle plasmid (red circles) exhibits a quasi-discontinuous jump to the high state whereas the pTAK102 control plasmid (blue triangles) exhibits a sigmoidal induction curve. Point 1 is taken from separate experiments measuring the high state of pTAK117 with no inducer. Points 3a and 3b are the high and low modes of a bimodally distributed cell population. The bimodality occurs due to natural fluctuations in gene expression and the close proximity of the toggle switch to its bifurcation point. Theoretical curves are calculated from equation (1) with the term $u/(1 + [IPTG]/K)^\eta$, where K is the dissociation constant of IPTG from LacR and η is the cooperativity of IPTG binding, replacing u in the denominator of equation (1b). The red curves show the stable steady states and the orange curve shows the unstable steady state of the toggle. The blue curve shows the steady state of the IPTG-inducible control plasmid. Model parameters for the theoretical curves are $\alpha_1 = 156.25$, $\alpha_2 = 15.6$, $\beta = 2.5$, $\gamma = 1$, $\eta = 2.0015$, $K = 2.9618 \times 10^{-5}$. **b**, Fraction of toggle cells in the high state at various concentrations of IPTG. The sudden switching to the high state is more clearly visible. High and low cell populations were divided as described for **c** below. **c**, Scatter plots (left plots) and histograms (right plots) illustrating the condition of the toggle cells at points 2, 3 and 4 (of **a**) near the bifurcation point. High-state and low-state cell populations are divided by the red line in the scatter plots. The two states of the toggle are clearly evident in the bimodally distributed cells (point 3).

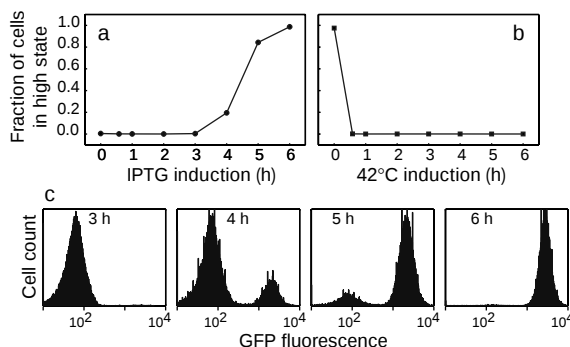


Figure 6 pTAK117 switching time. **a, b**, The fraction of cells in the high state is plotted as a function of the induction time. Cells were divided between high and low states as in Fig. 5c. **c**, Switching of pTAK117 cells from the low to the high state by IPTG induction. The cell population is illustrated at four time points. Cells begin switching between 3 and 4 h as shown by the appearance of a bimodal distribution. The switching is complete by 6 h.

architecture, rather than the engineering of proteins and other regulatory elements, to obtain desired behaviours. In addition, the reasonable agreement between the toggle theory and experiment indicates that the theoretical design of complex and practical gene networks is a realistic and achievable goal. Moreover, the genetic toggle switch exemplifies a forward engineering approach to the study of gene regulation in which synthetic gene circuits serve as highly simplified, highly controlled models of natural gene networks. As a practical device, the toggle switch, which requires only transient rather than sustained induction, may find applications in gene therapy and biotechnology. Finally, as a cellular memory unit, the toggle forms the basis for 'genetic applets'—self-contained, programmable, synthetic gene circuits for the control of cell function. □

Methods

Numerics

All theoretical curves were calculated numerically from equation (1) (Box 1) using Matlab (Mathworks), XPP-AUTO, software for simulation and analysis of differential equations (G. B. Ermentrout, University of Pittsburgh, available at <http://www.pitt.edu/~phase/>), or AUTO, a bifurcation package included in the XPP-AUTO software (E. Doedel, McGill University).

Plasmid construction

Plasmids were constructed using basic molecular cloning techniques as described in standard cloning manuals^{22,23}. Restriction enzymes were from New England Biolabs and Promega; PfuTurbo polymerase was from Stratagene; all other enzymes were from New England Biolabs; all synthetic oligonucleotides were from Operon Technologies. All genes, promoters and transcription terminators were obtained by PCR amplification using PfuTurbo proofreading polymerase and synthetic primers with overhanging ends containing the appropriate restriction sites. Ribosome binding sites were included in the overhanging ends of the primers. Site mutations were performed using either Stratagene QuickChange or ExSite.

Genes, promoters and transcription terminators were obtained as follows: P_{trc}-2 from pTrc99a (AP Biotech); P_L from pXC46 (ATCC); p_{tetO}-1 by total synthesis according to the published sequence²⁰; lacI from pTrc99a; clts from pGW7 (ATCC); tetR from pCDNA6/TR (Invitrogen); gfpuv from pGFPuv (Clontech); gfpmut3 from pJBA111 (gift of J. B. Andersen, Technical University of Denmark); and rrnT1T2 terminators from pTrc99a. All plasmids contained the ampicillin resistance region and ColE1 and origin of replication from the pTrc99a plasmid. All cloning was performed by TSS transformation²² into *E. coli* strain JM2.300 (CGSC), JC158 (CGSC) or TAP106 (ATCC).

DNA sequencing was performed using a Perkin-Elmer ABI Prism 377 sequencer.

Strains, growth conditions and chemicals

The host cell for all promoter assays and toggle switch experiments was *E. coli* strain JM2.300 (λ -, lacI22 rpsL135 (StrR), thi-1) (CGSC strain 5002). JM2.300, which contains few mutations, is a fast-growing strain that can tolerate enormous overexpression of plasmid-bound genes. Because JM2.300 contains no λ repressor and carries a nonfunctional Lac repressor (lacI22), it is an ideal host for the toggle switch.

All cells were grown in LB medium (Difco) with 100 μ g ml⁻¹ ampicillin plus inducers as

indicated in the text. All Type I and pIKE series plasmids were grown at 37 $\pm$ 1 °C, unless otherwise indicated. All pTAK series plasmids were grown at 32 $\pm$ 1 °C except during thermal induction. Thermal induction was carried out at 42 $\pm$ 1 °C, unless otherwise indicated. For all expression tests, cells were maintained in logarithmic growth phase by periodic 500–1,000-fold dilution into fresh medium.

Ampicillin and IPTG were from Sigma. Anhydrotetracycline was from ACROS Organics. All other chemicals were from Fisher.

Assay of gene expression

All expression data were collected using a Becton–Dickinson FACSCalibur flow cytometer with a 488-nm argon excitation laser and a 515–545-nm emission filter. Before assay, cells were pelleted and resuspended in 0.22 μ m filtered PBS (58 mM Na₂HPO₄, 17 mM NaH₂PO₄, 68 mM NaCl, pH = 7.4). Cells were assayed at low flow rate and fluorescence was calibrated using InSpect green fluorescent beads (Molecular Probes). All measurements of gene expression were obtained from three independent cultures maintained simultaneously under identical conditions. For each culture, 40,000 events were collected. All flow data were converted to ASCII format using MFI (E. Martz, University of Massachusetts, Amherst, available at <http://marlin.bio.umass.edu/mcbfacs/flowcat.html#mfi>) and analysed with Matlab.

Received 15 September; accepted 23 November 1999.

- Modod, J. & Jacob, F. General conclusions: teleonomic mechanisms in cellular metabolism, growth and differentiation. *Cold Spring Harb. Symp. Quant. Biol.* **26**, 389–401 (1961).
- Ptashne, M. *A Genetic Switch: Phage λ and Higher Organisms* (Cell, Cambridge, Massachusetts, 1992).
- Ishiura, M. *et al.* Expression of a gene cluster *kaiABC* as a circadian feedback process in cyanobacteria. *Science* **281**, 1519–1523 (1998).
- Schellenberger, W., Eschrich, K. & Hofmann, E. Self-organization of a glycolytic reconstituted enzyme system: alternate stable stationary states, hysteretic transitions and stabilization of the energy charge. *Adv. Enzyme Regul.* **19**, 257–284 (1980).
- Glass, L. & Kauffman, S. A. The logical analysis of continuous, non-linear biochemical control networks. *J. Theor. Biol.* **39**, 103–129 (1973).
- Glass, L. Classification of biological networks by their qualitative dynamics. *J. Theor. Biol.* **54**, 85–107 (1975).
- Glass, L. Combinatorial and topological methods in nonlinear chemical kinetics. *J. Chem. Phys.* **63**, 1325–1335 (1975).
- Kauffman, S. The large scale structure and dynamics of gene control circuits: an ensemble approach. *J. Theor. Biol.* **44**, 167–190 (1974).
- Thomas, R. Logical analysis of systems comprising feedback loops. *J. Theor. Biol.* **73**, 631–656 (1978).
- Thomas, R. Regulatory networks seen as asynchronous automata: a logical description. *J. Theor. Biol.* **153**, 1–23 (1991).
- Tchuraev, R. N. A new method for the analysis of the dynamics of the molecular genetic control systems. I. Description of the method of generalized threshold models. *J. Theor. Biol.* **151**, 71–87 (1991).
- Arkin, A. & Ross, J. Computational functions in biochemical reaction networks. *Biophys. J.* **67**, 560–578 (1994).
- Bhalla, U. S. & Iyengar, R. Emergent properties of networks of biological signaling pathways. *Science* **283**, 381–387 (1999).
- Yagilo, G. & Yagil, E. On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophys. J.* **11**, 11–27 (1971).
- Shea, M. A. & Ackers, G. K. The O_R control system of bacteriophage Lambda: a physical-chemical model for gene regulation. *J. Mol. Biol.* **181**, 211–230 (1985).
- Smith, T. F., Sadler, J. R. & Goad, W. Statistical-mechanical modeling of a regulatory protein: the Lactose repressor. *Math. Biosci.* **36**, 61–86 (1977).
- Arkin, A., Ross, J. & McAdams, H. H. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected *Escherichia coli* cells. *Genetics* **149**, 1633–1648 (1998).
- McAdams, H. H. & Arkin, A. Stochastic mechanisms in gene expression. *Proc. Natl Acad. Sci. USA* **94**, 814–819 (1997).
- McAdams, H. H. & Arkin, A. Stimulation of prokaryotic genetic circuits. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 199–224 (1998).
- Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I₁-I₂ regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
- Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).
- Ausubel, F. M. *et al.* *Current Protocols in Molecular Biology* (Wiley, New York, 1987).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Plainview, New York, 1989).
- Edelstein-Keshet, L. *Mathematical Models in Biology* (McGraw-Hill, New York, 1988).
- Kaplan, D. & Glass, L. *Understanding Nonlinear Dynamics* (Springer, New York, 1995).
- Yagil, E. & Yagil, G. On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophys. J.* **11**, 11–27 (1971).
- Rubinstein, S. I. *Introduction to Mathematical Biology* (Wiley, New York, 1975).

Supplementary information is available at *Nature's* World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank M. Bittensky and T. Yoshida for providing access to their flow cytometer; Y. Yu for his suggestions on plasmid construction; C. Sabanayagam for his technical advice; and C. Chow for his mathematical advice. This work was supported by the Office of Naval Research and the College of Engineering at Boston University.

Correspondence and requests for materials should be addressed to J.J.C. (e-mail: jjcollins@bu.edu). Plasmid sequences are available at <http://cbd.bu.edu/abl/toggle>.

Sensors, safety and SPR

Biosensors, biochips, bioeverything.

Biosensor technology

Remedios www.remedios.uk.com

Using bugs to counter pollution

The Remedios company was launched in November 1999 to develop biological sensor technology for the diagnosis and clean-up of polluted land. The company's standard suite of biosensors assesses the overall toxicity of environmental samples such as soil, sludge, sediment, groundwater, waste material and industrial effluent. The biosensors, developed at the University of Aberdeen, are based on a spectrum of bacterial strains representative of those found in the environment, and those that would be used in any bioremediation process. The biosensor organisms, containing genes from bioluminescent marine bacteria, respond to a toxic challenge by reducing light output. Remedios are marketing this technology in a series of screening and clean-up services, called Toxmonitor, Toxmap, Interprotox, Restore and Troubleshooter.

Reader Service No. 100

OX 100

Walden Precision Apparatus Ltd

www.wpaltd.co.uk

Measure dissolved oxygen in the field

This new DO meter from Walden features a robust galvanic cell allowing rapid *in situ* calibration on air. This is combined with automatic temperature compensation and a simultaneous temperature display to increase accuracy and reliability. The probe in the OX 100 can be used in a minimal flow of water and can automatically compensate for salinity and barometric pressure with simple keypad input. No polarization time is required, giving rapid results with this easy-to-use instrument. To aid use in the field, a 16-dataset, non-volatile memory can store results for recall.

Reader Service No. 101



Air show: WPA's meter and probe.



Glove box: keep your distance with 4H liners.

Safety 4H liners

Sentinel Laboratories

Safety and comfort in dry-box isolators

The use of dry-box isolators is common in pharmaceutical applications, but dry-box gloves may remain in place for many months, raising problems of hygiene. And as the choice of material for such gloves is limited, there may be some chemicals for which the gloves do not provide protection. Sentinel's Safety 4H dry-box glove liners counter these problems. The liners have a laminate construction, are thin (2.7 mm) and resistant to chlorinated solvents, hydrocarbons, ketones and many solvent mixtures. e-mail: senlab@global.co.uk

Reader Service No. 102

RDE-2

BAS Technicol Ltd www.bioanalytical.com

An electrode with biosensor applications

For laboratories interested in hydrodynamic electrochemical techniques, the RDE-2 from BAS is a rotating-disk electrode with excellent acceleration and deceleration capabilities. In addition to stationary electrode experiments, applications include biosensors, catalysts, semiconductors and redox characterization. The system comes complete with cell vials and a range of electrodes. A waterjacketed cell is an option. Built-in gas control allows purging or blanketing of the sample prior to or during experiments and rotation speeds from 50 r.p.m. to 10,000 r.p.m. can be adjusted either remotely or manually.

Reader Service No. 103

Cytosensor

Molecular Devices Corp. www.moldev.com

A microphysiometer system detects functional responses from living cells in minutes

The Cytosensor system evaluates cellular responses by measuring small changes in

extracellular acidification. As energy metabolism is coupled to cellular ATP usage, any event that perturbs cellular ATP levels, such as receptor activation and initiation of signal transduction, will cause a change in acid excretion. In addition to ATP usage, receptor activation may also directly affect acidification by altering the Na^+/H^+ exchange. An example of the system in action is in the measurement of the functional expression of the human chemokine receptor ChemR13, using a patented biosensor to measure changes in cell metabolism, without the need for prior knowledge of the signal-transduction pathway.

Reader Service No. 104

inoLab

WTW Measurement Systems

www.wtw-inc.com

Compact electrochemical measurement

With the inoLab series of benchtop electrochemical instruments, bench space is organized through a multifunction box that serves as an electrode stand, storage compartment for cables, and a calibration station complete with beakers, calibration standards, and storage. Available parameters include pH/OPP ISE, conductivity, DO, temperature and multiparameter configurations. The meters are available in three functional levels, from level 1 for basic routine measurements, to level 3 for research-grade performance.

Reader Service No. 105

Äkta prime

Amersham Pharmacia www.apbiotech.com

Simple protein purification at all stages of the purification process

Typical applications for this addition to the Äkta design chromatography platform include fusion-protein purification,



Prime mover: chromatography from AP.

new on the market

antibody purification and buffer exchange. Äkta prime gives push-button protein purification. Components include a high-quality system pump and a combined UV, pH and conductivity monitor. Other Äkta systems are tailored for development of fast purification methods and routine peptide and nucleic-acid analysis.

Reader Service No. 106

Biacore 3000

Biacore www.biacore.com
High-performance affinity-based biosensor systems

The main applications for Biacore 3000 will be in biomolecular binding measurements. BIATechnology detects binding events between two or more molecules, in real time, without the use of labels. Molecules do not need to be purified. The technology relies on the phenomenon of surface plasmon resonance (SPR) which occurs when surface plasmon waves are excited at a metal/liquid interface. Light is directed at, and reflected from, the side of the surface not in contact with sample, and SPR causes a reduction in the reflected light intensity at a specific combination of angle and wavelength. Biomolecular binding events cause changes in the refractive index at the surface layer, detected as changes in the SPR signal. In general, the refractive index change for a given change of mass concentration at the surface layer is practically the same for all proteins and peptides, and is similar for glycoproteins, lipids and nucleic acids.

Reader Service No. 107

FreeZone

Labconco Corporation www.labconco.com
A compact and customizable benchtop freeze dryer

Labconco's FreeZone 6-Liter Benchtop Freeze Dry System is intended for moderate



Upright model: freeze-drying on the bench.

sample loads and when the lab has limited floor space. The system can be customized to specific research requirements with accessories including several drying chambers, manifold and a stoppering tray dryer. The system features a user-friendly control panel providing touch-pad switches for controlling refrigeration, vacuum and LCD menu selection. Automatic one-button start-up can be used to initiate the collector cool-down and vacuum pull-down sequences. Refrigeration and vacuum may also be controlled manually.

Reader Service No. 108

IASys Auto+ Advantage

Affinity Sensors www.affinity-sensors.com
Top-of-the-range fully automated biosensor

The IASys Auto+ Advantage is designed for the medium- to high-throughput laboratory requiring complete automation of data collection and analysis functions. The Auto+ Advantage uses the same resonant mirror detection system as is used in the Auto+ and integrates this with a robotic system, Peltier thermostated dual-sample stage and robot-accessible buffer and reagent racks. Manual operation is possible via Point and Transfer (PAT) software, a windows-based graphical user interface — alternatively, real-time, unattended via Adaptive Sequence Control (ASK) software.

Reader Service No. 109

GeneTac LS IV

Genomic Solutions
www.genomicsolutions.com

High-throughput laser-based biochip analyser

The GeneTAC LS IV Biochip Analyzer is a laser-based, high-throughput biochip imager and analyser. It is fitted with three internal lasers and one external laser to detect CY3, CY5, FITC and Texas Red fluorescent labels. The spatial resolution is fully adjustable from 2 to 30 μm . The system includes the GeneTAC Integrator, a fully integrated analysis package that uses a powerful database



Change TAC: GeneTAC analysis on a chip.

(Microsoft SQL Server 7.0) for effective queries and data portability.

Reader Service No. 110

Dissolved oxygen probes

Sentek <http://members.aol.com/SENTEKuk/>
Two polygraphic dissolved-oxygen probes for laboratory measurements

These two new probes, Types 601 and 602, are designed for laboratory measurements and both have a standard 12-mm body diameter. In air-saturated water (20.9% oxygen), one has an output current of 600 nA (Type 601), the other 60 nA (Type 602). The residual current in zero-oxygen solutions for both electrodes is less than 1% of the current in saturated water. One or two ATC sensors (Thermistor or Pt100/1000) may be fitted to customers' specific requirements. The electrode is supplied complete with 2 membrane assemblies and a 50-ml bottle of filling solution. A 50-ml bottle of Zero Oxygen solution is also available for calibration purposes.

Reader Service No. 111

DNeasy

Qiagen www.qiagen.com
Harvesting plant DNA

The DNeasy Plant Maxi Kit allows rapid purification of total cellular DNA from up to 1 g of plant tissue. High-quality DNA is isolated from a wide range of plant species and tissue types, including troublesome sources rich in polysaccharides, polyphenols and other secondary metabolites. Typical yields of 30 μg (*Arabidopsis*) to 260 μg (wheat) of DNA are obtained depending on genome size and ploidy. Samples may be fresh, frozen or dried. The easy spin procedure provides rapid homogenization of lysed samples and efficient removal of cell debris using the unique QIAshredder. Pure DNA (genomic, chloroplast and mitochondrial), free from contaminants and enzyme inhibitors, is isolated ready for use in downstream applications such as PCR and Southern blotting.

Reader Service No. 112

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card.

BACs
YACs
PACs
cDNAs

Research Genetics provides clones and colony membranes for a number of human, mouse, and rat libraries as well as a custom screening service for the libraries.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

Check out our homepage at <http://www.resgen.com>

READER ENQUIRY NO. 6